Juuso Taskinen

Functional genomics of ORP2 or protrudin knockdown in human umbilical vein endothelial cells

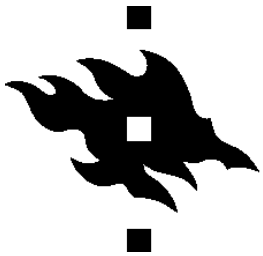
Master's programme in genetics and molecular bioscience

Faculty of Biological and Environmental Sciences

University of Helsinki

Master's thesis

2019



HELSINGIN YLIOPISTO

HELSINGFORS UNIVERSITET

UNIVERSITY OF HELSINKI

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Faculty Faculty of Biological and Environmental Sciences	Degree Program Master's programme in Genetics and Molecular bioscience

Author
Juuso Taskinen

Title
Functional genomics of ORP2 or protrudin knockdown in human umbilical vein endothelial cells

Study track
Molecular and analytical health biosciences

Level	Month and year	Number of pages
Master's Thesis	August 2019	55 + 2 appendices

Abstract

Human umbilical vein endothelial cells are responsible for maintaining and forming new vessels from existing ones, in a biological process called sprouting angiogenesis. Sprouting angiogenesis is a crucial mechanism for the resolution of hypoxia and normal development of tissues. It also plays a key role in internal plague hemorrhages, which can lead to embolisms and other cardiovascular complications. Angiogenesis is also crucial for cancer development. Sprouting angiogenesis is initiated by hypoxic tissue excreted vascular endothelial growth factor gradient, which induces normal endothelial cells into either a proliferative stalk cell or a signal sensing tip cell phenotype. Both of these cell types depend on the rapid flow of lipids to their plasma membrane, either to form plasma membrane protrusions in tip cells or as new plasma membrane material in dividing stalk cells. This flow is envisioned to involve both vesicle-mediated and non-vesicular mechanisms. A major non-vesicular route of lipid transfer occurs at membrane contact sites via lipid transport proteins. Furthermore, lipids can be transported to the plasma membrane by the direct fusion of vesicles or endosomes with the plasma membrane

This thesis set out to explore the role of two membrane contact site proteins, oxysterol-binding protein- related protein 2 and protrudin, in angiogenesis and lipid transfer. Their role was examined by RNA-sequencing transient knock-down samples of these proteins in HUVECs. The RNA-sequencing data was examined by differential expression, gene ontology overrepresentation and gene set enrichment analyses. Gene expression analysis provided almost 10 000 significantly changed transcripts (adjusted p-values < 0.05), in each silenced cell type. The distribution of differentially expressed genes in oxysterol-binding protein- related protein 2 silenced cells, is skewed toward negative fold changes, whereas the distribution of differentially expressed genes in protrudin silenced samples is normally distributed. The results also show significant changes in gene ontologies related to proliferation, cell cycle, angiogenesis as well as hypoxia in both sample types. Gene set enrichment analysis showed upregulation in angiogenesis related pathways, such as the PI3K-Akt and MAPK pathways, in both samples. Significant downregulation was present in cell cycle related pathways and cholesterol biosynthesis pathway in both ORP2 and protrudin silenced samples.

Keywords
RNA-sequencing, ORP2, protrudin, membrane contacts site, angiogenesis, endothelial cell, gene expression, functional genomics.

Supervisor or supervisors
Vesa Olkkonen Ph.D, Prof.; Amita Arora Ph.D

Where deposited
E-thesis

Additional information

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiivistelmä

Tiedekunta	Koulutusohjelma
Bio- ja ympäristötieteellinen tiedekunta	Genetiikan ja molekulaaristen biotieteiden maisteri ohjelma

Tekijä
Juuso Taskinen

Työn nimi
ORP2 ja protrudiini hiljennyksen funktionaalinen genetiikka, ihmisen napanuoran suonen soluissa

Opintosuunta
Molekulaariset ja analyyttiset bioterveystieteet

Työn laji	Aika	Sivumäärä
Pro gradu	Elokuu 2019	55 + 2 liitettä

Tiivistelmä

Ihmisen napanuoran suonen solujen primaari tarkoitus on huolehtia suonien kunnosta ja muodostaa uusia suonia vanhoista prosessissa, jota kutsutaan angiogeneesiksi. Angiogeneesin päätarkoitus on estää kudoksien ravinteiden sekä hapen puute ja se on kriittinen osatekijä kudosten normaalissa kasvussa. Angiogeneesillä on myös ratkaiseva rooli ateroskleroottiset plakkien sisäisissä verenvuodoissa, jotka saattavat johtaa embolioiden sekä muiden kardiovaskulaaristen komplikaatioiden muodostumiseen. Angiogeneesillä on myös ratkaiseva rooli monien syöpien kehityksessä. Uusien suonten muodostumisen laukaisee hapen puutteesta kärsivän kudoksen synnyttämä endoteelikasvutekijä gradientti. Tämä gradientti indusoi normaaleja endoteelisoluja muuttamaan fenotyyppinsä, joko jakautuviksi juurisoluiksi tai signaaleja aistiviksi kärkisoluiksi. Molemmat näistä solutyypeistä ovat riippuvaisia lipidien nopeasta virtauksesta solukalvolle, mitä ne käyttävät muodostaakseen solukalvo protruusioita tai uuden solukalvon rakenteeksi. Tämä lipidien virtaus voi tapahtua joko vesikelivälitteisen kuljetuksen tai ei-vesikelivälitteisen kuljetuksen avulla. Tärkeä ei-vesikelivälitteinen kuljetuksen muoto tapahtuu kalvokontaktipaikoilla, jossa lipidiensiirtoproteiinit siirtävät lipidejä kalvojen välillä. Lipidejä voidaan siirtää solukalvolle myös endosomien ja vesikelien fuusiolla.

Tämän pro gradu -työn tarkoitus oli tutkia kahden kalvokontaktiproteiinin, oksysterolia sitovien proteiinien kaltaisen proteiini 2 sekä protrudiinin, roolia angiogeneesissä ja lipidien kuljetuksessa. Näiden kahden proteiinin shRNA-hiljennyksen vaikutusta HUVEC-soluissa tutkittiin RNA-sekvensoinnin avulla. RNA-sekvensointi dataa analysoitiin geeni-ilmentymis-, geeniontologia ylirepresentaation- sekä geenijoukko rikastuma-analyyseilla. Geeni-ilmentymisanalyysi paljasti noin 10000 differentiaalisesti ilmentynyttä transkriptiä, joiden korjattu p-arvo oli alle 0.05. Oksysterolia sitovien proteiinien kaltaisen proteiini 2 hiljennetyissä näytteissä geenien ilmentymisen jakauma on vinoutunut ali-ilmentymisen puolelle, kun taas protrudiini hiljennetyissä näytteissä geeninen ilmentyminen oli normaalisti jakautunut. Geeniontologia ylirepresentaationanalyysin mukaan geeniontologiat, jotka liittyvät solujen jakautumiseen, solusykliin, angiogeneesiin sekä hapen puutteeseen, ovat tilastollisesti merkitsevästi yliedustettuja. Geenijoukko rikastuma-analyysi osoitti tilastollisesti merkittävää yli-ilmentymistä angiogeneesiin liittyvissä signalointireiteissä, kuten PI3K-Akt ja MAPK. Tilastollisesti merkittävää ali-ilmentymistä esiintyi solusykliin liittyvissä signalointireiteissä sekä kolesteroli biosynteesi metaboliareitissä.

Avainsanat
Angiogeneesi, endoteelisolu, geenien ilmentyminen, ORP2, protrudiini, RNA-sekvensointi, kalvokontaktiproteiini, funktionaalinen genetiikka.

Ohjaaja tai ohjaajat
Vesa Olkkonen FT, prof.; Amita Arora FT

Säilytyspaikka
E-thesis

Muita tietoja

Table of Contents

Abbreviations

ARF	ADP ribosylation factor
Dll4	Delta like ligand 4
ER	Endoplasmic reticulum
EC	Endothelial cell
FFAT	Two phenylalanines in an acidic tract motif
GPL	Glycerophospholipid
HUVEC	Human umbilical vein endothelial cells
KD	Knockdown
LTP	Lipid transfer protein
MCS	Membrane contact site
OSBP	Oxysterol-binding protein
ORP	Oxysterol-binding protein- related protein
PH	Pleckstrin homology domain
PM	Plasma membrane
PtdCho	Phosphatidylcholine
PtdInsP	Phosphatidylinositol phosphate
RT-qPCR	Reverse transcriptase quantitative PCR
SNARE	soluble NSF (N-ethyl-maleimide-sensitive fusion protein) attachment protein receptor
VAMP	Vesicle associated membrane protein
VAP	Vesicle associated membrane protein- associated protein
VEGF	Vascular endothelial growth factor
VEGFR	Vascular endothelial growth factor receptor

1 Introduction

The aim of this thesis was to explore the gene expression profile of knockdown (KD) human umbilical vein endothelial cells (HUVEC) with reduced expression of protrudin and oxysterol-binding protein–related protein 2 (ORP2). Especially any changes in the biological functions and pathways related to angiogenesis and lipid metabolism were of interest. The effects of reduced expression of the aforementioned proteins were studied in basal conditions (normal ECGM 2 media, grown at 37 °C , 5% CO_2), in order to find any perturbations that are due to the transient silencing of these proteins. The expression profile was characterized with RNA-sequencing data that were processed in a standard workflow, which is explained in the methods section. The count matrix produced from this workflow was controlled for shRNA off-target effects by a prediction algorithm, and any downregulated genes flagged by it were removed from the analysis. Further analysis performed on the RNA-sequencing data followed a standard workflow: Exploratory, differential gene expression, gene ontology overrepresentation and finally gene set enrichment analysis. This workflow was chosen because it is the most logical in terms of finding meaningful changes in biological functions. Exploratory data analysis is done to quality control the data and gene ontology analysis was performed to see, which biological functions have changed. Gene set enrichment analysis is the final step to explore, which molecular functions have changed within the altered biological functions. This thesis will follow a familiar flow of topics. First an introduction to the main themes related to this thesis will be discussed. Then methods, as well as the results will be explained in detail, and the final sections of this thesis includes a discussion on the significance of the results and brief account of further experiments that could be performed.

2 Background

This section explains and elucidates the biological processes in which ORP2 and protrudin are involved in. First, endothelial cells (EC) and their intrinsic biological functions are discussed, in order to paint a bigger picture of the biological processes, where ORP2

and protrudin are of vital importance. Mainly their significance in regulation of cellular membranes during angiogenesis, which is one of the most crucial functions of ECs. After these sections, this thesis will dwell deeper into the molecular mechanisms, through which ORP2 and protrudin might be responsible for the control of plasma membrane (PM) modulation and therefore angiogenesis.

2.1 Endothelial cells

ECs line the walls of blood and lymph vessels as a thin monolayer of cells. Their main function is to act as a barrier between the liquid and cells within the lumen and rest of the body, as well as the trafficking of molecules from the lumen to tissues. The most basic viable structure that ECs can form is a combination of ECs, basal lamina and pericytes. However, this structure can extend to more complex features, such as the combination of smooth muscles attached to the basal lamina, which are able to control the diameter of the vessel. ECs are also responsible for extending the vessel network of both blood and lymph vessels during development. This vessel network formation is called angiogenesis and without it tissues would die of hypoxia and malnutrition (Alberts, Johnson, Lewis, Morgan, & Raff, 2014, p 1235 - 1238).

2.2 Angiogenesis

Since the hypothesis of this thesis is that ORP2 and Protrudin might play crucial roles in the modulation of the plasma membrane of HUVECs during angiogenesis, it is important to understand how HUVECs migrate to tissues and how angiogenesis is initiated. To this end the next chapter explores the mechanism, regulation and complications of angiogenesis.

Usually angiogenesis is defined as the formation of new blood vessels, but it actually refers to the sprouting of new vessels from existing ones. The induction and formation of new vessel sprouts are dependent on the reaction of differentiated ECs to chemotactic stimulus. These tip cells are induced by a vascular endothelial growth factor (VEGF) gradient, which along with other angiogenic signals, guides them to invade tissues. The

VEGF gradient is triggered in tissues by the lack of nutrients or oxygen, and the transcription of VEGF is induced by increased levels intracellular hypoxia-inducible factor 1α (Geudens & Gerhardt, 2011). The selection of which ECs turn into tip cells depends on how well they are able to bind VEGF to their VEGF receptors (VEGFR), as well as the expression level of the delta-like protein 4 (Dll4). The binding of VEGF to the VEGFR2 begins a signaling process that represses neighbouring cells and stops them from becoming tip cells. This process turns the neighbouring cells into stalk cells through lateral inhibition that is mediated by the Dll4/Notch signalling pathway (Geudens & Gerhardt, 2011). Stalk cells are proliferating ECs and their main function is to elongate the angiogenetic sprout, as well as form the blood vessel lumen.

The aforementioned signaling process acts through the signalling pathway from VEGF to Notch, which is as follows: VEGF binding to VEGFR2 induces expression of Dll4, which in turn binds to Notch. The Notch intracellular domain is cleaved after Dll4 binding, and it initiates the Notch signalling pathway. This pathway downregulates the expression of *VEGFR2*, *VEGFR3*, *NRP1*, *Dll4* and *CXCR* in the neighbouring cells, which inhibits them from becoming tip cells (Geudens & Gerhardt, 2011). This inhibition also leads to the upregulation of VEGFR1 and jagged1. VEGFR1 competes with the downregulated VEGFR2 for VEGF, which in turn reduces the expression of Dll4 in stalk cells. Jagged1 acts as a defective notch ligand since it does not activate the notch pathway, this in turn leads to the competitive binding of jagged1 to the tip cells' notch receptors, inhibiting notch activation, therefore keeping it in its transient tip cell phenotype. The whole process of vessel sprouting and the feedback process between tip and stalk cells can be seen in figure 1.

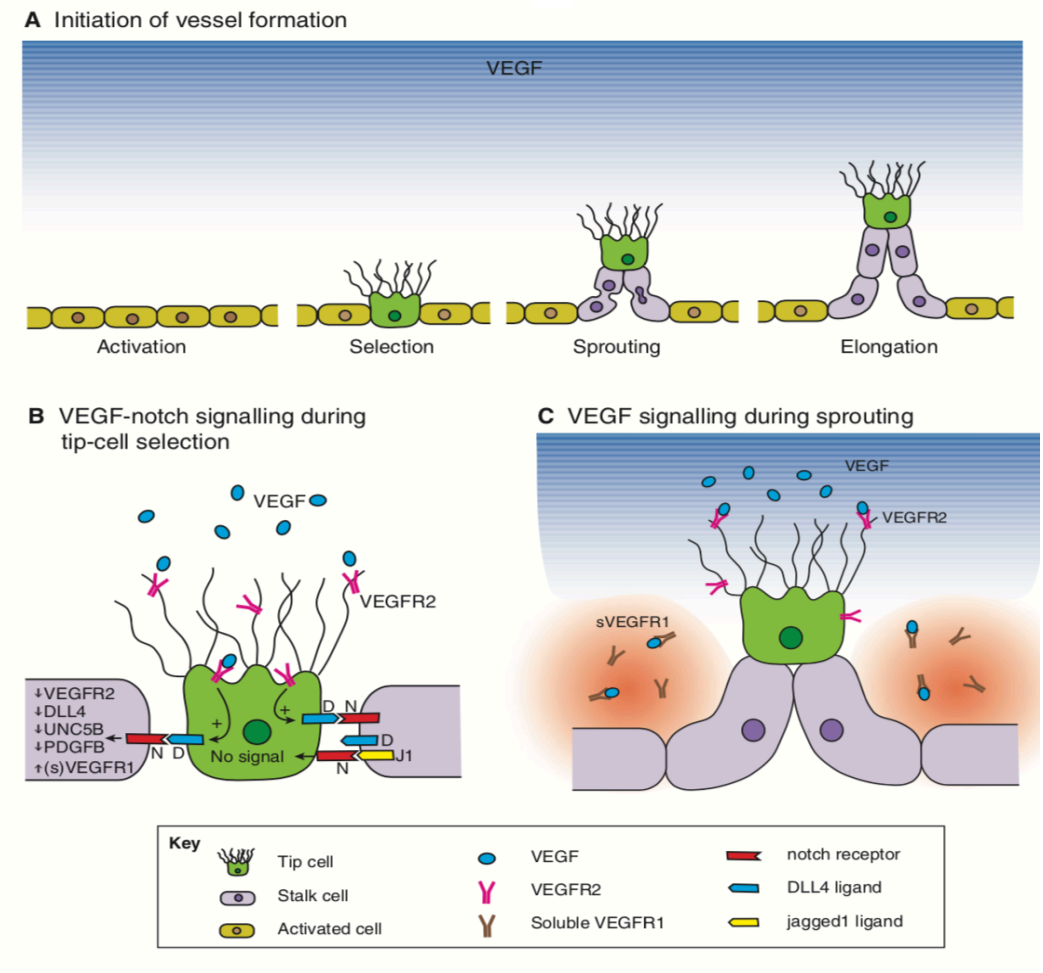


Figure 1. A) VEGF gradient activates ECs (yellow) and primes them for sprouting. Out of the activated ECs only a few are selected as tip cells (green) due to VEGF/notch regulation. Lateral inhibition induces stalk cell (purple) phenotype in the tip cells' neighboring cells, beginning vessel sprouting. Elongation of the sprout happens through the combination of tip cell guidance by angiogenetic signals and proliferation of stalk cells. B) Activation of VEGFR2 by VEGF begins the regulatory feedback loop, which keeps both tip and stalk cells in their respective transient phenotypes. VEGFR2 induced expression of the notch ligand Dll4 begins the activation of the notch pathway in stalk cells. This reduces the expression of VEGFR2 and Dll4 in stalk cells, which makes them less sensitive to VEGF. This in turn reduces the chances of the notch signaling pathway activating in other cells. Furthermore, the expression of other angiogenic signaling receptors (PDGFB, UNC5B) is downregulated and the expression of VEGF decoy receptors (VEGFR1) and defective notch ligands (Jagged1) is upregulated in stalk cells, which further cements their phenotype. C) Soluble VEGFR1 further increases the VEGF gradient around the tip cell by binding to VEGF next to the tip cell and therefore decreasing the VEGF concentration in the lateral sides of the tip cell (adapted from Geudens & Gerhardt, 2011).

The aforementioned feedback loop keeps both cells in their respective transient phenotypes until the VEGF gradient disappears, i.e. the hypoxia and nutrient deficiency in the tissue is resolved (Geudens & Gerhardt, 2011).

2.2.1 Guidance of sprouting angiogenesis

New blood vessels cannot grow aimlessly toward the source of the VEGF gradient; therefore, a mechanism is needed to steer the new blood vessel to the correct direction. This guidance is the main responsibility of the tip cell that reacts to both the aforementioned VEGF gradient, as well as other signals from the tissue, which are sensed through lamellipodia and filopodia. Lamellipodia are extending PM sheets driven by a dynamic mesh-like actin structure, whereas filopodia are narrow PM protrusions with a core of filamentous actin. Both filopodia and lamellipodia are able to probe the cellular environment, and their response to environmental cues is either the depolymerisation or polymerisation of the actin filaments. This leads to the respective extension or retraction of the PM protrusion, depending on the environmental cue (Geudens & Gerhardt, 2011). Filopodia and lamellipodia can attach to the extracellular matrix by forming focal contacts. These attachment points allow cells to fix themselves to the extracellular matrix, which help them to migrate within tissues. Cell movement occurs with the contraction of intracellular actin microfilaments called stress fibres, which drag the cell toward focal contacts. Since cells are attached to the extracellular matrix, tip cells have to degrade it in order to invade tissues. The degradation of the extracellular matrix is aided by matrix metalloproteinases, which are expressed by tip cells. Metalloproteinases also regulate angiogenesis and are essential for the formation of the invading vessel in adults. It has also been demonstrated that membrane type-1 metalloproteinase is downregulated in stalk cells (Geudens & Gerhardt, 2011).

2.2.2 Sprout elongation and lumen formation

The elongation of the sprouting vessel happens through the proliferation of stalk cells and so far, two models have been proposed as the mechanism of elongation. One is the push model, where stalk cells are thought to push the tip cell deeper into tissues as they divide. Second is the pull model, which explains the elongation of the vessel by the tip cell pulling stalk cells with it, as it invades tissues. The pull model has gained more support over the years and studies have shown that the tip cell indeed exerts a pulling force on the stalk cells (Geudens & Gerhardt, 2011). In support of the push model, studies have also revealed that angiogenetic sprouts regress when stalk cell proliferation is decreased,

suggesting that the stalk cells are also pushing the tip cell forward (Geudens & Gerhardt, 2011). This evidence would imply that a combination of the push and pull models could be responsible for the elongation of sprouts.

A new sprout cannot function as a blood vessel without a lumen and how the vessel lumen is formed is still debated. However, there have been a few suggestions as to the mechanism of lumen formation during sprout elongation, which are usually referred to as: the intracellular vacuole coalescence, intracellular vacuole exocytosis and luminal repulsion models, which are illustrated in figure 2. Intracellular vacuole coalescence is described as the fusion of intracellular vacuoles within the cells. The fused vacuoles generate ever larger void within the cell until similar voids within other cells fuse together and create the lumen (Geudens & Gerhardt, 2011). Intercellular vacuole exocytosis explains lumen formation by the fusion of exocytotic vacuoles that make new intercellular space by fusing together outside the cells, eventually making enough space for a lumen to form between two cells (Geudens & Gerhardt, 2011). Lumenal repulsion is a mechanism where negatively charged CD34-sialomucins locate to VE-Cadherin sites between two cells. This creates enough electrostatic repulsion between the two cell membranes that VE-cadherin molecules relocate to the lateral side of the cells, and the membranes withdraw from their original positions. This essentially creates an intercellular void between the two cells, which then fuses with others like it in neighbouring cells, creating the lumen (Geudens & Gerhardt, 2011).

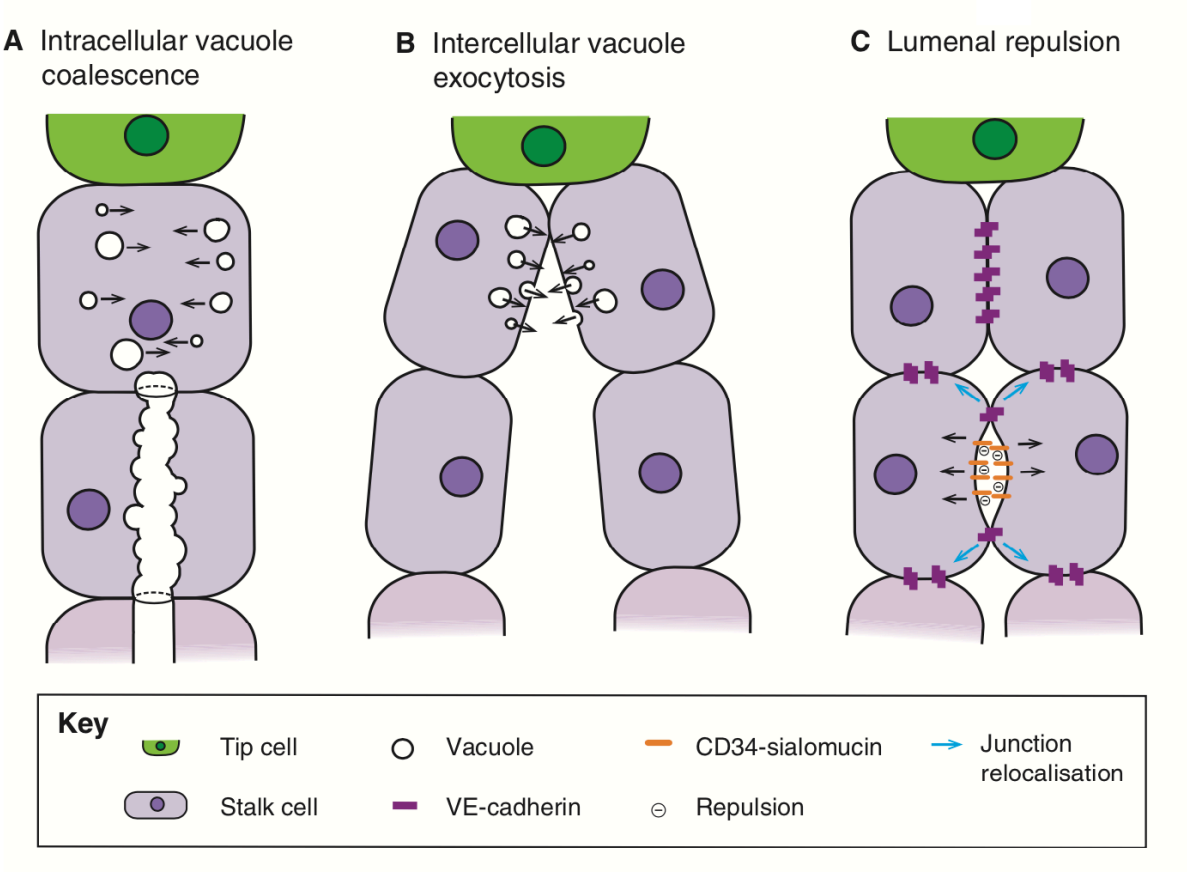


Figure 2. Different mechanisms of lumen formation by endothelial stalk cells during sprouting angiogenesis. A: Intracellular vacuole coalescence explains the lumen formation by the action of intracellular vacuoles fusing together forming the lumen within the cell. B: Intercellular vacuole exocytosis, explains lumen formation by the fusing events of exocytotic vesicles, which form the intercellular space C: Lumenal repulsion, explains lumen formation by the electrostatic repulsion (dot) of CD34-sialomucin molecules (orange), which leads to the depression of the apical membrane and relocation of VE-cadherin junctions (purple) (Adapted from Geudens & Gerhardt, 2011).

The above examination of different aspects relating to angiogenesis is far from complete and much was left undiscussed. Angiogenesis is a hot research topic and some mechanisms related to it still remain a mystery. For example; Stalk cells continue proliferating even during lumen formation and how they are able to regulate the balance between tight EC junction and proliferation, in order to form a sealed and unobstructed lumen, requires more elucidation (Geudens & Gerhardt, 2011).

2.2.3 Angiogenesis in disease

Normal angiogenesis is crucial for a multitude of different physiological functions, such as wound healing and development, and it is vital that the effects of angiogenesis are

also reversible. Angiogenesis is a complex mechanism and therefore there are many points where perturbations can lead to deleterious effects. For example; continuous feedback process between tip and stalk cells, even in the absence of a VEGF gradient, can cause unwanted angiogenesis. Tip cells reacting to wrong cues might also steer the blood vessel to wrong direction, which can lead to vessels growing in wrong tissues at the wrong time. Some of these angiogenesis related diseases will be discussed in the following chapter.

Cancer is probably one of the most common contexts where disturbed angiogenesis is discussed. This is probably due to the fact that one of the hallmarks of cancer is the tumours' ability to initiate angiogenesis, and many modern cancer treatments try to perturb this. Angiogenesis is crucial for the survival of the tumour, since it needs oxygen and nutrients for rapid growth, as well as a route for spreading into other tissues i.e. metastasis. One of the treatments used for cancers is the anti-VEGF antibody bevacizumab, which is one of many FDA approved chemotherapies. Bevacizumab is used to treat advanced metastatic cancer, such as non-small cell lung cancer, renal cell cancer and colorectal cancers. For other cancers, for example hepatocellular carcinoma or thyroid cancer, VEGF receptor tyrosine kinase inhibitors are more commonly used (Moreno, Purushothaman, & Purushothaman, 2012; Potente, Gerhardt, & Carmeliet, 2011). Unfortunately, cancers can circumvent VEGF stimulated sprouting angiogenesis in many ways: VEGF-independent vessel growth mediated through other proangiogenic signals and sprouting angiogenesis independent vessel growth mediated by, for instance vascular mimicry (Potente et al., 2011). Anti-angiogenesis treatment can also lead to severe complications, such as hypertension, thrombosis or embolisms, and therefore it might not be a viable treatment for cancer patients with atherosclerosis (Moreno et al., 2012)

Angiogenesis also has an important role in cardiovascular diseases, which with their many complications, are one of the leading causes of morbidity and mortality in the world (Moreno et al., 2012). In particular, atherosclerotic plaque formation initiates angiogenesis through hypoxia, which is caused by increased oxygen demand through inflammation and reduced oxygen supply, due to decreased blood flow by the thickening of vessel walls (Moreno et al., 2012). This can lead malformed capillaries growing within

the plaque, which in turn cause intraplaque haemorrhage. Intraplaque haemorrhage causes the plaque to become unstable, which can lead to pieces tearing from the plaque or the whole plaque breaking apart, eventually leading to infarctions. (Moreno et al., 2012). Un-regressing angiogenesis can also pose problems to tissues, which have altered mechanical properties due to expanding vessel networks. ECs that are constantly stimulated by angiogenetic factors will keep growing vessel networks and in addition these networks keep expanding due to sprouting angiogenesis. This leads to a large network of small vessels that can rupture or collapse during mechanical stress leading to problems, such as heart failure (Moreno et al., 2012).

2.3 Membrane contact sites

Angiogenesis is heavily reliant on the timely and accurate traffic of lipids to the PM of tip and stalk cells, which need them for extension of PM protrusions or proliferation and lumen formation. Non-vesicle mediated facilitators of this lipid flow are membrane contact site (MCS) proteins, which regulate the transfer of lipids within cells, but these proteins have not so far been researched in ECs. The existence of MCSs has been known for years, and they were first described in the late 1950s and early 1960s (Copeland, 1959; Porter, 1957; Robertson, 1960). Their role in biological functions has been a mystery, but some clarity on their function has emerged in the recent decades, which will be discussed in the following chapters.

MCSs are usually described as membrane areas where two organelles, such as the endoplasmic reticulum (ER) and PM, are tethered together in close contact. However, not every instance where two membranes are in interaction can be described as a MCS. Four distinct criteria define an MCS (Prinz, 2014):

- Two membranous organelles are in close contact (15–30 nm apart) through tethering proteins, which facilitate the trafficking of material between organelles
- Membranes are in close proximity, but they do not fuse together at MCSs
- Certain proteins as well as lipids are in abundance at MCSs

- The function and organisation of one, or both of the organelles is affected.

The scope of this thesis will not allow the complete examination of different MCSs and trafficking of different molecules between different organelles, since the number of permutations is high. Therefore, the following sections will concentrate on discussing MCSs between ER, Golgi and endosomes, as well as the trafficking of lipids between them. Many areas, such as calcium exchange between ER, mitochondria and endosomes, will not be discussed and neither will the lipid exchange between ER and mitochondria.

2.3.1 Tethering

Tethering MCS proteins and protein complexes have an important function, not only in tethering two membranes together, but also in the maintenance of the MCSs. MCS proteins that act as tethers can be identified by three main functions: MCS proteins can mediate tethering independently, MCS proteins are regulated by feedback processes and MCS proteins can perform multiple functions (Prinz, 2014). MCSs are mostly maintained by several different proteins, for which an excellent example is a *Saccharomyces cerevisiae* ER to PM junction. Studies have demonstrated that at least six of the ER proteins need to be removed, in order to drastically reduce the tethering between ER and PM at the aforementioned junction. This junction is serviced by calcium- and lipid-binding proteins, such as Tcb1-3 and Ist2, as well as Scs2 and Scs22, which are the yeast orthologues of vesicle-associated membrane protein-associated proteins (VAP). This suggests that proteins at MCSs can mediate tethering independently of each other (Prinz, 2014). PM-ER tethering through VAPs can, for example, be achieved by oxysterol-binding protein (OSBP)- related proteins (ORP), which mediate the tether through their two phenylalanines in an acidic tract motif (FFAT) and pleckstrin homology (PH) domain. The PM-ER tethering is achieved by the VAP binding to the FFAT motif of ORP and, in turn, PH domain of the ORP binding to phosphatidylinositol phosphates (PtdInsP) in the PM (Prinz, 2014). Tethering at MCSs is also a regulated process. This fact was demonstrated by Mesmin *et al.*, (Mesmin et al., 2013) when they proved the negative feedback regulation of ER-Golgi OSBP tethers. Mesmin et al. demonstrated that when PtdIns(4)P is enriched at Golgi, OSBP binds to it and begins transferring PtdIns(4)P to

ER, where Sac1 phosphatase mediated hydrolysis of PtdIns(4)P prevents it from moving back to Golgi. Once the levels of PtdIns(4)P drop at the Golgi, the tethering between Golgi and ER is reduced due to the dissociation of OSBP from Golgi. Tethering proteins and complexes seem to also have other functions than just tethering membranes. For example, mitofusin-2 is primarily responsible for mitochondrial fusion, but it also functions as an ER–mitochondrial tethering protein (Prinz, 2014).

2.3.2 Membrane identity

Many tethering MCS proteins are able to identify the membrane they interact with a multitude of different protein domains as described earlier. This identification process is crucial, because without it MCS proteins would not be able to perform their different functions correctly, including lipid transport. Cellular machinery identifies membranes through a combination of both specific lipids and proteins that are found in membranes. Hallmark protein families that give membranes their identity, have so far been recognized to be the Rab, ARF (ADP ribosylation factor) and SNARE (soluble NSF (N-ethyl-maleimide-sensitive fusion protein) attachment protein receptor) protein families. The Rab protein family consists of proteins that are able to switch between an active GTP-bound form and an inactive GDP-bound form, which is catalysed by GDP/GTP exchange factor (GEF), as well as GTPase-activating protein (GAP). The Rab family of proteins act as regulators of vesicular tethering, trafficking and fusion, as well as the movement of cellular structures (Stenmark & Olkkonen, 2001). Like Rabs, the ARF protein family is a part of Ras superfamily of GTP-binding proteins, which localize to cellular membranes, such as Arf1 to the Golgi or Arf6, which localizes to the PM and endosomes. The main function of this protein family is to recruit, for example, vesicle coat proteins and lipid modifying enzymes to the membranes, which in turn modulate the membrane composition and shape (Donaldson & Honda, 2005). SNARE proteins are responsible for the fusion of vesicles to membranes. These proteins include homologues of SNAP-25, syntaxin and vesicle associated membrane proteins (VAMP), which mediate vesicle fusion to membranes. The suggested mechanism of vesicle fusion involves zippering of SNAREs, which can be summed up as the intertwining event of membrane bound SNAP-25 and syntaxin coiling together with a vesicle’s VAMPs. This sequence of events forces

the two membranes together, fusing them to one another (Chen, Scheller, & Medical, 2001).

PtdInsP have been shown to interact and attract proteins that have a phosphoinositide-interacting domain, which localize them to PtdInsP containing membranes. Since PtdInsP can be phosphorylated in a number of positions in their inositol group, as many as seven different PtdInsP can be formed. This diversity of phosphorylated inositol groups can be used to differentiate membranes based on their PtdInsP populations. Composition of PtdInsP on the membrane leads to the diversification of lipid binding-proteins that bind to the membrane, essentially creating an identity for the membrane with the combination of lipids and lipid bound proteins. Furthermore, the Rab, ARF and SNARE protein families are able to regulate the production, localisation and dynamics of PtdInsPs, which in turn affect the localisation and activity of these proteins (Hammond et al., 2012; Jean & Kiger, 2012; Kutateladze, 2010; Lemmon, 2008; Platre & Jaillais, 2016; M. L. A. Simon et al., 2016). Even though PtdInsPs are crucial for many signalling events that govern a host of cellular activities, such as membrane trafficking, migration and signalling transduction, only trace amounts can be found in cellular membranes (Di Paolo & De Camilli, 2006).

2.3.3 Protrudin and MCS proteins in endosome translocation

Protrudin is one of many MCS proteins that are responsible for the regulation of endosomal translocation in the cell. These proteins are able to identify and tether their target membranes and endosomes through the proteins and lipids mentioned in the previous section. Protrudin belongs to the FYVE-finger family of proteins, which are mostly responsible for the localisation of endosomes. Protrudin is an ER transmembrane protein that interacts with VAPs, and it is also capable of binding to late endosome Rab7 as well as Rab11. Neurite outgrowth is highly dependent on protrudin that functions as the mediator of Rab11 regulated membrane recycling. This membrane recycling of endosomes is critical for the quick and timely modulation of protruding membranes, such as the neurite growth cone. Protrudin also has been shown to have a role in other neurite-like protrusion forming biological processes in non-neuronal cells. Overexpression experiments of protrudin have proven that protrudin can induce kidney repair by initiating tubulogenesis (Zhang *et al.,* 2017). Protrudin might also have a role in other biological

processes that depend on protrusion generation in cells, such as angiogenesis. Protrudin mediates endosomal translocation by binding to Rab7 via its low complexity region (LCR) and PtdIns(3)P, which is prevalent on endosomal membranes, through its FYVE domain. Protrudin regulates late endosome localisation by recruiting kinesin-1 to the MCS and transferring it to the late endosome membrane localized coiled-coil domain containing protein FYCO1. This protein is localized to membranes through Rab7 and PtdIns(3)P, which tether it to late endosomes. The main function of FYCO1 is to act as a link between the kinesin-1 and late endosomes. Linking of kinesin-1 via FYCO1 to late endosomes promotes their transport to cell periphery (Raiborg, Wenzel, Pedersen, & Stenmark, 2016). Figure 3 illustrates the proposed mechanism of protrudin mediated late endosome translocation during neurite outgrowth. The localisation of endosomes, to and from the cell's periphery, is therefore regulated by ER–endosome MCS tethering, which affects kinesin and dynein motor proteins. Moving the tether also moves both organelles, which can be seen as branching of ER tubules from the rough ER that are associated with endosomes and almost devoid of ribosomes. Almost half of early endosomes (53%) and nearly all late endosomes (99%) are tethered to the ER while they are trafficked in the cytosol. This suggests that the number of MCSs increases as the endosomes mature (Wijdeven, Jongsma, Neefjes, & Berlin, 2015). The composition of ER–late endosome MCSs is also in flux during the maturation process and various proteins that regulate the localisation of late endosomes can be seen at these sites.

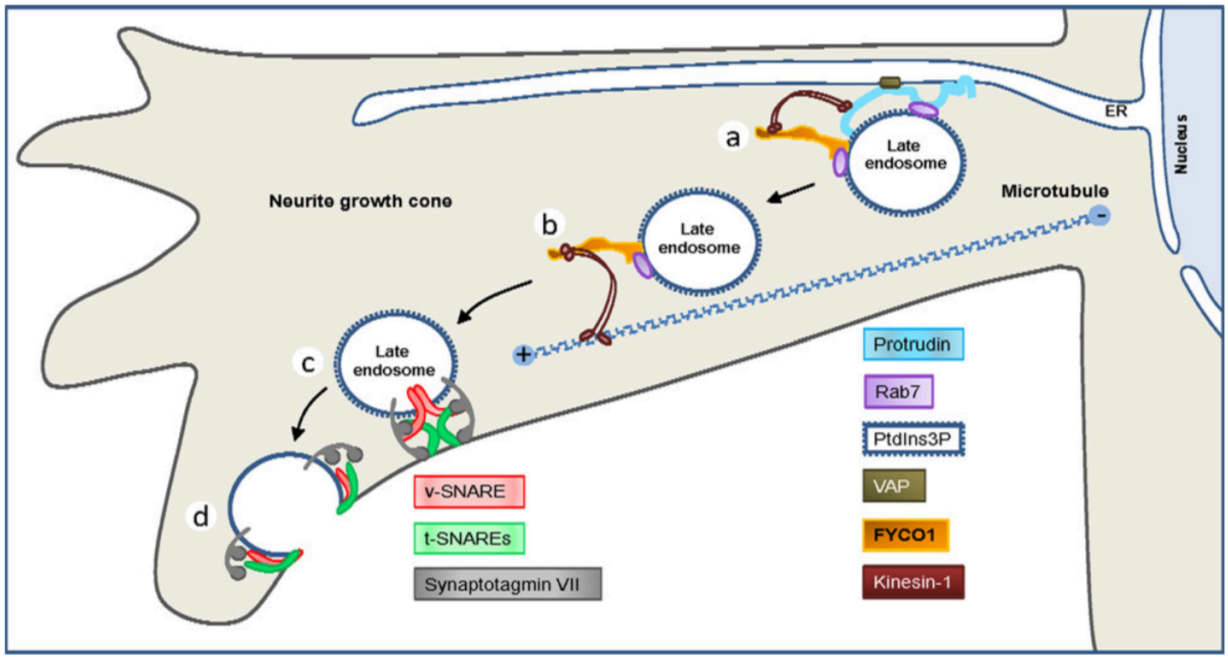


Figure 3. Protrudin mediated late endosome translocation in neurite growth cone. A: ER late endosome contact is established through protrudin PtdIns3P Rab7-GTP tether. Tethering leads to the recruitment of Kinesin-1, which binds to the late endosomal FYCO1. B: Kinesin-1 moves the late endosome toward the tip of the growth cone along microtubules. C and D: Late endosome fuses to the growth cone through vesicle fusion (adapted from Raiborg et al., 2016).

The accumulation of late endosomes at the cell periphery or at the centre of the cell has been proposed to be dependent on cholesterol levels (Wijdeven et al., 2015). Only a few mechanisms describing cholesterol-dependent localisation of late endosomes have so far been proposed. One of these mechanisms is dependent on ORP1L mediated cholesterol sensing on the surface of late endosomes. ORP1L changes conformation when its ORD domain binds cholesterol, which allows it to form a complex with the Rab7 GTPase, RILP (RAB-interacting lysosomal protein), the HOPS late endosome tethering complex and the dynactin-dynein motor complex. The formation of this complex results in the accumulation of late endosomes at the centre of the cell, mediated by minus-end trafficking along microtubules. In low cholesterol conditions, the ORD domain is not able to bind cholesterol and the conformation of ORP1L does not therefore change. The unchanged ORP1L conformation, in turn, keeps the FFAT motif free, and ORP1L binds to VAP instead of the aforementioned complex. FFAT motif binding to VAP then results in the reduced dynein facilitated minus-ended trafficking of late endosomes (Wijdeven et al., 2015).

Another protein that could be in charge of regulating the localisation of late endosomes, is the START domain-containing protein 3 (STARD3). Since STARD3 is capable of binding cholesterol with the START domain and it also contains a FFAT motif, it is very likely that it might have a role in sensing sterol at ER–late endosome MCSs. Some evidence that STARD3 is indeed associated with the localisation of late endosomes, has been obtained from KD experiments. These experiments have shown that STARD3 KD results in the trafficking of late endosomes to the cell periphery, as well as the reduction of actin patches on late endosomes. Furthermore, overexpression of STARD3 increases actin patches on late endosomes and locates them to the perinuclear region of the cell. The molecular machinery interacting with STARD3 is still unknown, but late endosomes seem to be associated to the actin skeleton through STARD3 (Alpy et al., 2013; Hölttä-Vuori, Alpy, Tanhuanpää, & Jokitalo, 2005).

2.3.4 Membrane lipids, transport between membranes and ORPs

This thesis has touched upon most of the four main themes related to membrane contact sites, but how and what kinds of lipids are trafficked between membranes, still remains undiscussed. To this end, the following chapter will examine the diversity of lipids within the cell, then move on to explain how lipids are transported between membranes and finally discuss the role of ORPs in this mechanism.

Membrane lipids

The diversity of different lipids in eukaryotic cells is wide, and over 1000 different lipid species with varying functions and structures have been identified. Lipids function as the building blocks of cellular membranes and make up around 50% of the total membrane mass (Alberts *et al.*, 2014, p 566). Lipids also act as an energy supply and storage, or as important factors in both signalling and intracellular trafficking. Lipids can be divided into eight distinct categories based on their structure and metabolic pathway of their production: fatty acids, glycerolipids, glycerophospholipids (GPLs), sphingolipids, sterol lipids, prenol lipids, saccharolipids and polyketides. The transport and metabolism of lipids is governed by roughly 5% of protein-coding DNA, and the most abundant lipid species are GPLs. They make up roughly 75 mol% of all cellular lipids, whereas the

second and third most abundant lipids are, sterol lipids at 12–14 mol% and sphingolipids at 8–12 mol%, respectively (Drin, 2014; van Meer, Voelker, & Feigenson, 2008). GPLs are so abundant because they are vital to the integrity of many membranes, which is evident from the fact that most intracellular membranes are composed of 40–44% of phosphatidylcholine (PtdCho). PtdCho is also vital for the PM of which 10–28 % is made up of PtdCho (Alberts et al., 2014, p 571). Sterols are also very important membrane lipids, cholesterol being one of the most abundant PM lipids making up around 17 - 23% of lipids in the PM. The abundance of different lipids in different membranes is illustrated in table 1.

Table 1.The amount of some lipids in different membranes (adapted from Alberts *et.al*, 2014, p 571)

Lipid composition of different membranes						
	Proportion of lipids in % (w/w)					
Lipid	Liver cell PM	Red blood cell PM	Myelin	Mitochondrion	ER	*E.coli*
Cholesterol	17	23	22	3	6	0
Phosphatidylethanolamine	7	18	15	28	17	70
Phosphatidylserine	4	7	9	2	5	trace
Phosphatidylcholine	24	17	10	44	40	0
Sphingomyelin	19	18	8	0	5	0
Glycolipids	7	3	28	trace	trace	0
Other	22	14	8	23	27	30

In addition to GPLs and sterols, sphingolipids, such as sphingomyelin and glycosphingolipids, can also be found in most eukaryotic membranes. The proportions of GPLs and sphingolipids are one factor that determines the rigidity of membranes. For example, membranes that have more sphingolipids are more gel-like, due to the fact that sphingolipids are able to pack more tightly. Whereas membranes, which have more GPLs, are more fluid compared to membranes that are abundant with sphingolipids, since GPLs are not able to pack so tightly (van Meer et al., 2008). Cholesterol can reduce the permeability of membranes to small water-soluble molecules; it prevents the hydrophobic tails of phospholipids from crystallizing and immobilizes fatty acyl chains in the surrounding region, making the membrane more rigid (Alberts et al., 2014, p 571).

Lipid transport between membranes and ORPs

The transfer of most of the aforementioned lipids is facilitated by lipid transport proteins (LTP), which are a group of proteins that bind and traffic lipids between membranes, or from membranes to proteins. LTPs can be divided into at least 10 protein families according to many factors, one of them being their lipid transfer domain. Even though LTPs are structurally highly diverse, they still share common features and means of lipid transfer. All LTPs have a highly hydrophilic surface and a nonpolar amphiphilic tubular pocket in which the transferred lipids are placed. They also share a lid structure that keep the bound lipids protected from an aqueous environment. Furthermore, LTPs share many similar localisation domains, such as the aforementioned PH domain or FFAT motif, which help LTPs to attach to different membranes. These domains are not just acting alone, but LTPs often have two different localisation domains, which in unison help them bind membranes to a close apposition, in order for them to facilitate lipid transfer between the membranes (Chiapparino, Maeda, Turei, Saez-Rodriguez, & Gavin, 2016). Figure 4 depicts 10 families of LTPs, their subcellular localisation and lipids they bind.

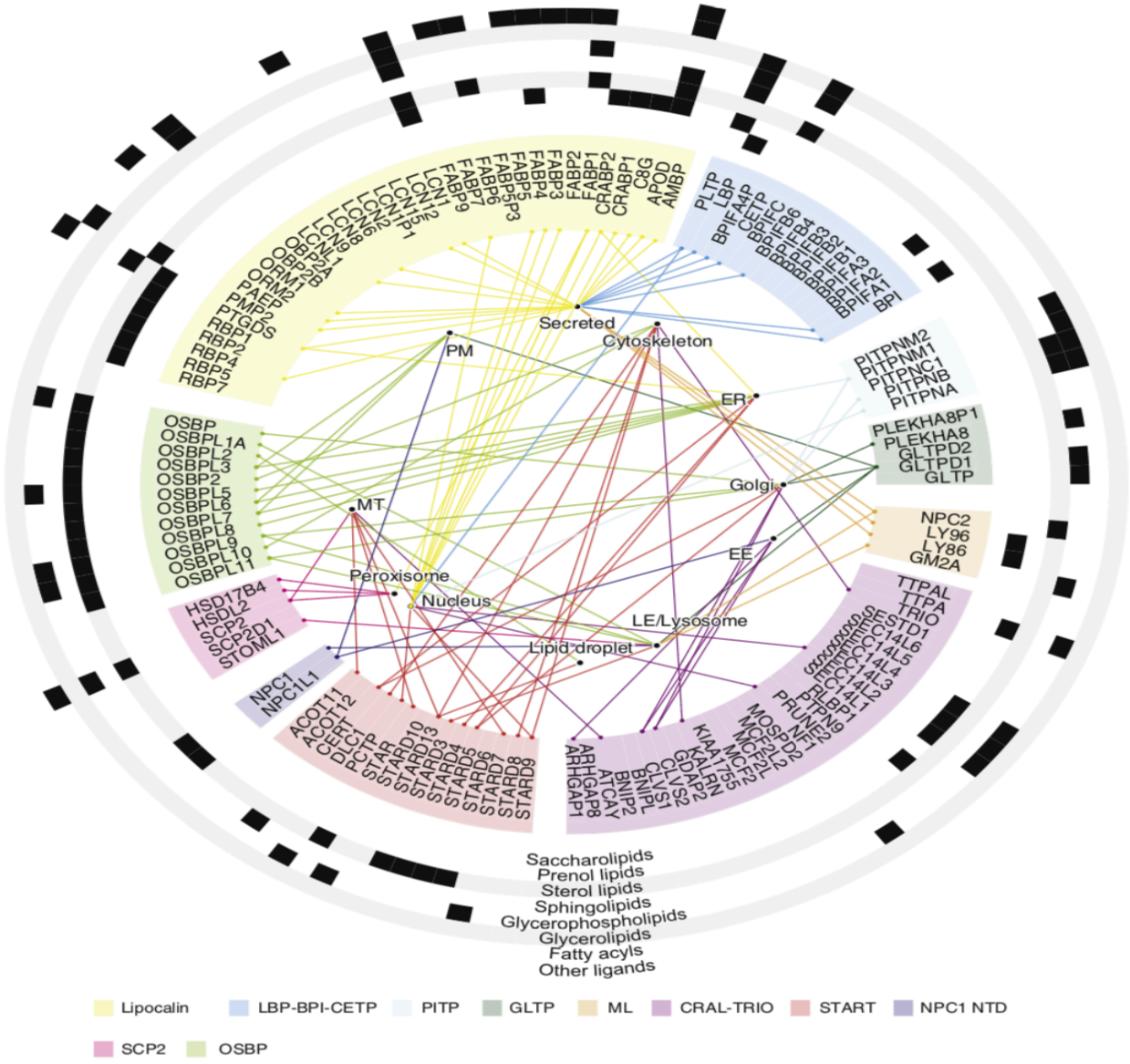


Figure 4. A chart describing ten different lipid transport proteins families in different colors (inner circle), their localization patterns (lines and dots) and different lipids species that the lipid transport proteins have affinity to (outer circles with black bars) (adapted from Chiapparino, Maeda, Turei, Saez-Rodriguez, & Gavin, 2016).

Lipid transfer between different organelles and endosomes can happen either through direct diffusion between two closely apposed membranes, vesicular trafficking or by LTP mediated transport. The ability of LTPs, to bind and traffic cholesterol and PtdInsPs between Golgi and ER was already mentioned earlier, but LTPs are capable of transferring a host of different lipids over a number of membrane permutations. An important lipid trafficked between ER and endosomes is cholesterol, which is integral to the rigidity, stability and permeability of membranes, and therefore efficient trafficking of it is essential to cells. The suggested mechanism of lipid transfer between membranes is

usually described as the following: First the LTP tethers two membranes together; an excellent example of this are some of the ORPs, which bind two membranes into close proximity with their FFAT and PH domains. After the membranes have been brought into close apposition, a lipid is extracted from the membrane by the LTP. After extraction, the lipid is inserted into the hydrophobic pocket within the protein, which shields it from the cytosol. The lipid is then transferred over the gap between two membranes by the protein, which finally inserts it to the outer membrane of the receiving membrane (Olkkonen, 2015). This mechanism is also illustrated in figure 5, which gives the outline of cholesterol traffic between Golgi and ER.

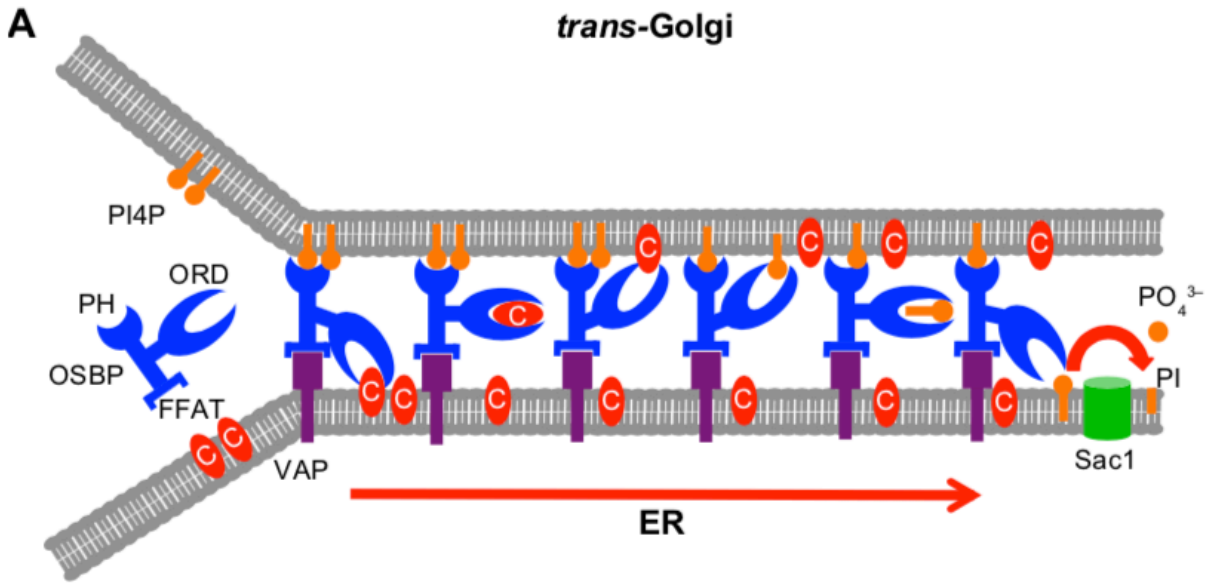


Figure 5. A model of ORP lipid transport mechanism at an ER – Golgi MCS. FFAT motif binds to VAP on the ER membrane and PH domain binds to PtdIns(4)P (PI4P) on the Golgi membrane. This action tethers the two membranes together, after which the transfer of cholesterol (C) can be mediated by the ORPs ORD domain. ORD extracts a cholesterol from the ER and ferries it over the membrane gap and places it in the Golgi membrane. ORD also extracts one PtdIns(4)P from Golgi to the ER Sac1 protein which then hydrolyses it to phosphatidylinositol. This cycle repeats until enough of PtdIns(4)P has been transported from the ER to Golgi and ORP disassociates from the Golgi membrane, which ends the transport of cholesterol against its concentration gradient (adapted from Olkkonen, 2015).

Whether or not the mechanism of cholesterol transport between ER, PM and endosomes is similar to the above described mechanism is still uncertain. However, there is a strong case to be argued that similar mechanism between these membranes does exist. Endosomes and PM both contain PtdIns(4)P and cholesterol, and VAPs exist on endosomes, which would make it likely that ORPs could be facilitating the transfer of

cholesterol between PM and endosomes. Studies have also shown that phosphatidylinositol-4,5-bisphosphate 4-phosphatase localizes to endosomal membrane, which acts similar to Sac1 on endosomes (Niebuhr et al., 2002).

ORP2 localises to the cytosol, lipid droplets and plasma membrane, and unlike number of other ORPs, ORP2 does not have a PH-domain so it is unable to tether its self to membranes with PtdIns(4)P. ORP2 has been shown to bind to the surfaces of lipid droplets and it can form tethering complexes with ER vesicle-associated membrane protein associated proteins that form tethers between ER and lipid droplets. Other evidence has shown that this complex is able to synthesise and inhibit the degradation of other lipids. Therefore, it would seem that the possible main function of ORP2 is to provide lipids and lipid precursors from lipid droplets to ER (Olkkonen, 2015). Two recent studies by Koponen *et.al.* and Wang *et.al*. showed that ORP2 is also able to control the subcellular localisation of cholesterol, especially its transport to the PM (Koponen et al., 2019; Wang et al., 2019). The latter study also provided evidence that ORP2 exchanges cholesterol to phosphatidylinositol-4,5-bisphosphate at the PM (Wang et al., 2019)

3 Aims

The aim of this master's thesis was to study the effects of ORP2 and protrudin KD in the biological functions of HUVECs through the exploration of RNA-sequencing data. The major hypothesis of this thesis was that KD of ORP2 or protrudin would hinder angiogenesis: In the case of ORP2 KD this might be due to hindered transport of cholesterol to the PM, due to the reduction of cholesterol traffic between endosomes and ER or endosomes and the PM. The protrudin KD might affect angiogenesis in a similar way that it impedes neurite outgrowths, i.e. the lack of endosomal transport. Other hypothesis includes; the lack of transport of other lipids to the plasma membrane due to protrudin KD and changes in lipid metabolism due to lack of lipid transport from lipid droplets and reduced control over cholesterol localization.

- First aim was to produce transient KD HUVECs of *OSBPL2* and *ZFYVE27* (i.e. protrudin) and sequence their transcriptome.

- Second aim was to process the base calls into high quality data, which could be used for downstream analysis.
- Third aim was to interpret the sequencing data through multiple different analysis steps and mine for the changes in biological functions that are due to the silencing of *OSBPL2* and *ZFYVE27,* especially any changes related to angiogenesis and lipid metabolism.

4 Materials and methods

The experimental work for this thesis was done according to the following flowchart (figure 6), which illustrates the order of the experiments and analyses performed.

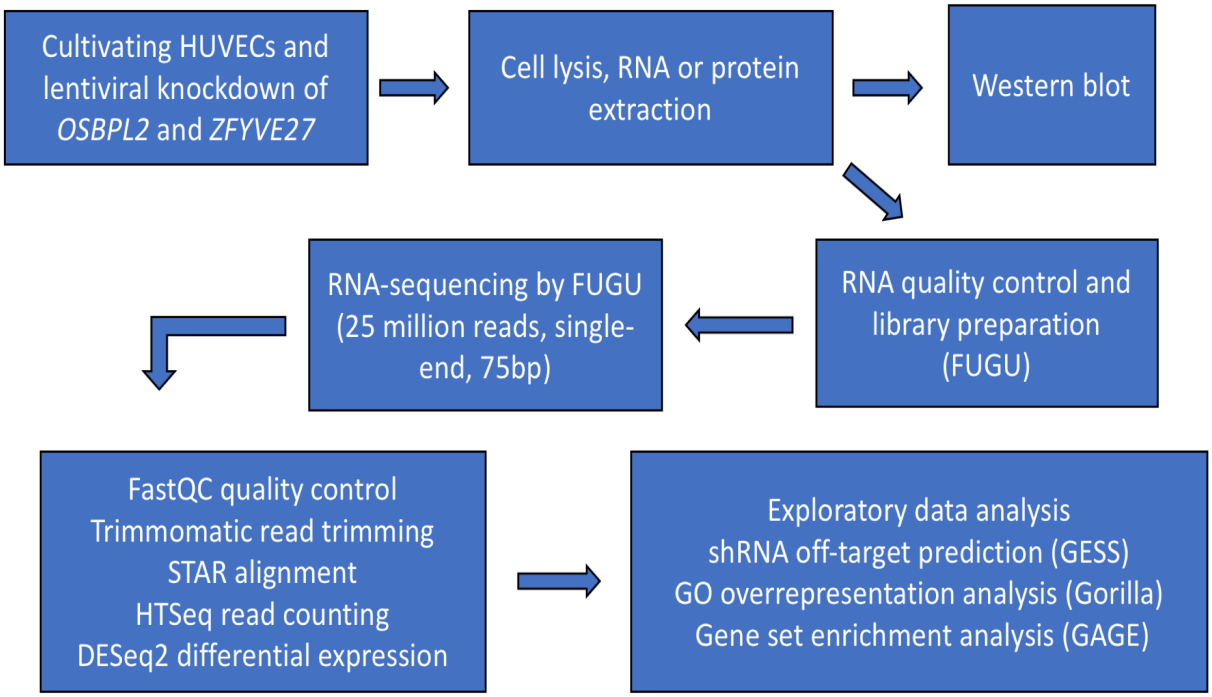


Figure 6. Flowchart outlining the experimental and analysis setup of the thesis

More detailed information about each experimental step are described under each guiding headline and the bioinformatic methods are mainly discussed as the programs used in the analysis of the data and details of the parameters used in the analysis are omitted, in order to streamline this section.

Cell culture

Primary cultures of HUVECs were purchased from PromoCell. Cell culture vessels for HUVECs were coated with a solution of fibronectin 10 µg / mL, gelatine 0.05% in 1X PBS for ½ – 2 hours and the solution was removed before any cells were seeded. HUVECs were grown in full ECGM2 media (PromoCell), trypsinization was done by washing the cells twice with 5 mL of warm 1X PBS and then adding 2 mL of Trypsin-EDTA (PromoCell) and coating the cells. After coating and a brief incubation, the Trypsin-EDTA was removed and cells were incubated at 37 °C, 5% CO_2 for 2 – 3 minutes. Cells were collected to a fresh 15 mL Falcon tube by adding 2 – 5 mL of media to the culture vessel and pipetting the media back and forth on the vessel surface, after which the cells were counted with a BioRad TC20 Automated cell counter. Collected cells were pelleted by centrifuging for 5 min at 800 RPM (Heraeus Megafuge 1.0). Media was removed, and cell pellets were resuspended in an appropriate amount of new media to achieve an optimal cell density for each experiment.

Lentiviral KD in HUVECs

Lentiviral particles were purchased from the University of Helsinki, FUGU core facility. Roughly $2.5*10^6$ cells/mL of passage 4 HUVECs were resuspended into ECGM2 media with double supplements. To 1 mL of cell suspension, 750 µL of fresh lentiviral stock was added. 1.75 mL of the transduction mix was added to each fibronectin – gelatine coated 6-well plate well and cells were incubated at 37 °C, 5% CO_2 for 48 hours. Puromycin selection of HUVECs was done by changing the transduction media to ECGM2 full media containing 2,5 µg/ml (stock 10 mg/ml, 1:4000-dilution) of puromycin. Cells were selected over night at 37 °C, 5% CO_2 and lysed next morning.

RNA extraction

RNA extraction was done according to the manufacturer's instructions using the Quiagen-RNAeasy mini 50 kit (Quiagen) with the following modifications: DNAse digest was used, centrifugation (Heraeus Biofuge pico) at steps 6 and 7 were done at 13000 rpm, and samples were eluted to 30 µL of RNase free water. Samples were stored at - 80 °C.

Cell lysis, protein concertation estimation and Western blot analysis

Lysis of cells for protein extractions was done by removing media, washing the cells with cold 1X PBS twice and then pipetting an appropriate amount of lysis buffer to the culture vessel. The composition of the lysis buffer can be found in table 3.

Table 2. Composition of the lysis buffer used for the extraction proteins from HUVECs

Lysis buffer composition for 10 ml of complete buffer		
Compound	Concentration	Volume
1X Lysis buffer stock - 10 mM HEPES pH 7.6 - 10% glycerol - 150 mM NaCl - 0,5 mM MgCl2		8,6 mL
0,5 % Triton X-100 (Sigma)	10%	0,5 mL
0,5 % sodium deoxycholate (Sigma)	10%	0,5 mL
1:25 PIC (Roche)	2X	0,4 mL

Cells were collected to a fresh tube and vortexed briefly before incubating on ice for 5 minutes. After, the tubes were centrifuged at 13000 rpm for 12 minutes at 4 °C, then cell lysate was removed to a new tube, and the sample was used for Western blotting.

Protein concentration of each sample measured done according to manufacturer's instructions (Thermo Fisher Scientific, Pierce BSA Protein Assay) using Thermo Fisher Scientific Multiskan FC equipment. Western blots were performed using 10% polyacrylamide gels that were made according to manufacturer's instructions (BioRad TGX Stain-Free FastCast GELS). Samples with total protein amount of 10 µg and 1X laemmil sample buffer were boiled at 100 °C for 5 minutes and loaded onto gels, which were run in 1X TGST running buffer at 300 V (BioRad, PowerPac HV) for roughly 25 – 30 minutes, until loading dye was either at the bottom of the gel or it had run through it. Gels were imaged before transfer and membranes after transfer using the BioRad ChemiDoc Touch Imaging System. Proteins were transferred from gels to PVDF membranes according to manufacturer's instructions, using the Trans-Blot Turbo RTA Transfer Kit (LF-PVDF) and Trans-Blot Turbo Transfer System (BioRad). After transfer the membranes were blocked at RT for 30 minutes using an appropriate blocking-buffer (TBST-5% milk, TBST-5% BSA etc.) After blocking membranes were washed three times for 5 minutes using 1X TBST. Washed membranes were incubated with the primary

antibody (ORP2 or Protrudin) over-night at 4 °C and after incubation the membrane was washed as described before. The membrane was then incubated with and appropriate secondary antibody (for example, 1:2000 HRP-anti-rabbit in TBST-5% milk) for 2 hours at RT. After the incubation the membranes were once again washed and then incubated with developing solution according to manufacturer's instructions (Thermo Fisher Scientific, Pierce ECL Western Blotting Substrate). Membranes were imaged every minute using the ChemiDoc, with a 5 minutes exposure.

Next generation RNA sequencing

Next generation RNA sequencing was performed by the University of Helsinki, functional genomics sequencing unit (FUGU). RNA quality control was done with the Tapestation (Agilent) according to manufacturer's instructions. Library preparation was done with the NEBNext Ultra Directional II kit after polyA purification. Library preparation quality control was also performed using the Tapestation. Sequencing was done with the Illumina NextSeq 500 sequencer (NextSeq High Output 1 x 75 bp).

Bioinformatics

Trimming, alignment, read counting and differential expression analyses were performed using the Chipster suite (Kallio et al., 2011). The following workflow was used to generate the differential expression data: FASTA files were quality controlled using FastQC (A. Simon, 2010) and trimmed using Trimmomatic (Bolger, Usadel, & Lohse, 2014). Single-end reads were aligned to the reference genome GRCh38, with STAR (Dobin et al., 2012). Mapped reads were counted using HTseq (Pyl, Anders, & Huber, 2014), with the ENSEMBL GRCh38 annotation file (Zerbino et al., 2018) as a reference. Differential expression analysis was performed using DESeq2, with a cut-off value for Benjamini-Hochberg adjusted p-values of 0.05 (Love, Huber, & Anders, 2014).

Short hairpin RNA off-target effects were verified by acquiring all 3' UTR sequences for each gene Ensemble ID from downregulated genes. The transcript sequences were annotated to Ensemble IDs with biomaRt (Brazma et al., 2005; Durinck, Spellman, Birney, & Huber, 2009) and then the reverse compliment of the seed sequence of each shRNA

(table 4) was referenced against the 3'UTR sequences to find a match, with an inhouse developed R script (appendix 1). Each shRNA seed sequence was assumed to work in a similar manner to miRNA seed sequences. The performance of the inhouse developed R script was also compared against a more robust shRNA off-target predictor, GESS (Sigoillot et al., 2012; Yilmazel et al., 2014)

Table 3. The shRNAs sequences used to silence ORP2 and protruding in HUVECs, seed sequence highlighted in yellow.

shRNA sequences used for ORP2 and protrudin (*ZFYVE27*) knockdown	
Gene	shRNA sequence
OSBPL2	CCGGCGGGAATGATTGACTTAGACACTCGAGTGTCTAAGTCAATCATTCCCGTTTTTTG
ZFYVE27	CCGGGTGTAACCAGACCTTGAGCAACTCGAGTTGCTCAAGGTCTGGTTACACTTTTTTG

Gene ontology overrepresentation analysis was performed using GOrilla (Eden, Lipson, Yogev, & Yakhini, 2007; Eden, Navon, Steinfeld, Lipson, & Yakhini, 2009). Gene set enrichment analysis was performed using either the Pathview web application, GAGE R package or consensuspathdb (Brouwer, Blanchard Jr, Luo, Pant, & Bhavnasi, 2017; Kamburov et al., 2011; Luo, Friedman, Shedden, Hankenson, & Woolf, 2009) and results were mapped onto KEGG-pathway graphs using Pathview (Brouwer & Luo, 2013; Kutmon et al., 2015; van Iersel et al., 2008) or Wikipathway maps using Pathvisio (Bohler et al., 2015; Pico et al., 2011).

5 Results

Western blot analysis, quality control and exploratory data analysis are discussed first. Exploratory data analysis is crucial, since if the samples cannot be separated from each other based on the gene expression, then downstream analysis of the data would not produce meaningful results. The shRNA target prediction is the next step in insuring that the data is of high quality and it is also needed to distinguish real downregulation from off-target effects. Gene ontology overrepresentation analysis is discussed next, and it is a convenient way to get biologically relevant information from the data. Not only does it compare your data against curated data sets, but it also produces intuitive plots, which help to narrow down biological functions that have been affected by the gene silencing.

Gene set enrichment analysis is the final step of the process and it can be used to find affected pathways within a biological function. Gene set enrichment analysis further narrows down the search for biologically meaningful changes due to the gene silencing. It is important to note that all the tools used to derive the results use probability in one way or another, which means that false discoveries are possible, and therefore some healthy scepticism should be practiced when interpreting the results.

Western blot and exploratory data analysis

Western blot analysis of ORP2 and protrudin lentiviral KD samples show that reduction of mRNA levels can also be seen at the protein level, within the time span of an ordinary KD experiment i.e. approximately 72 hours (figure 7). Levels of actin can be seen to be steady in each sample and therefore it is reasonable to conclude that the reduction in the intensity of the bands in the gel is due to the reduction in the protein level, and not due to sample loading error. Level of ORP2 has been reduced significantly more compared to control than the level of protrudin, but a reduction can be seen also for the latter.

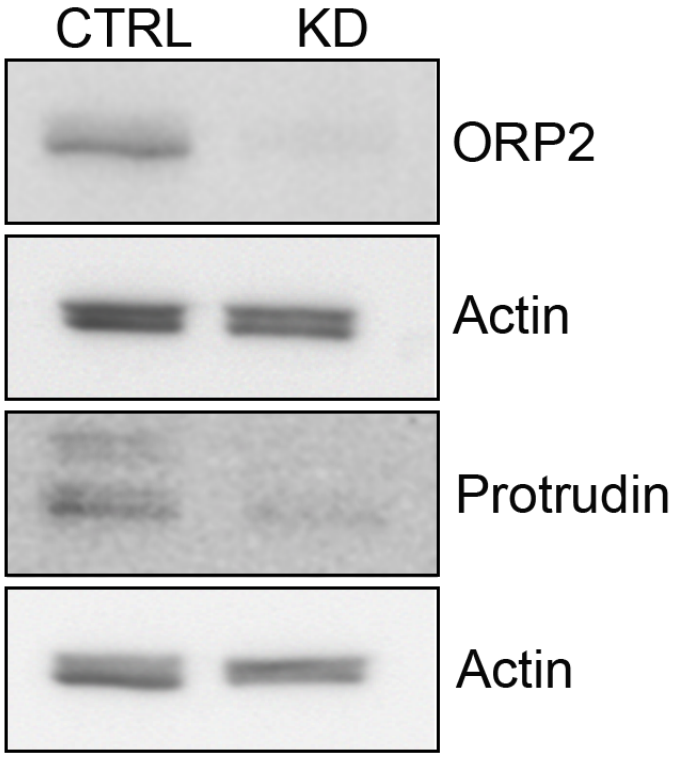


Figure 7. Levels of ORP2, protrudin and actin in both control and lentiviral KD samples. Data provided by Annika Koponen

Quality control plots from multiQC illustrate clearly that the sequence data produced by FUGU was of excellent quality (figure 8). Overwhelming majority of bases are above the phred score of 30 and the mean phred score in all base positions is above 30.

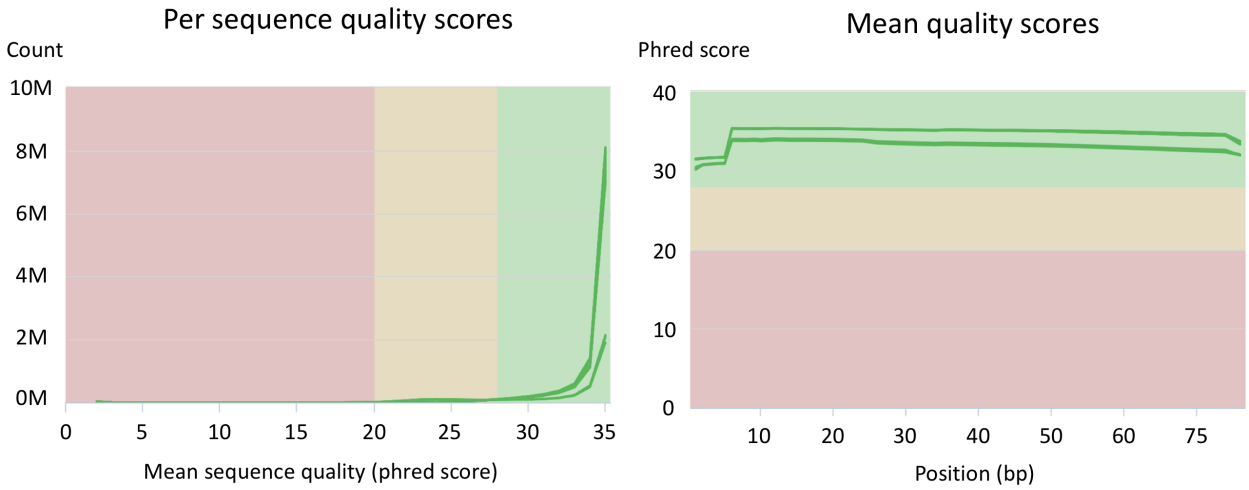


Figure 8. MultiQC report plots of the per sequence quality score and mean quality score of bases in all sequencing lanes (n = 3, technical replicates = 4, 4 lanes per replicate).

Other quality control metrics (table 5) also illustrate that the quality of the sequencing data is excellent. For example, mean sequences length is above 70, mean GC content is 51%, mean per base N content is less than 5% along the total length of the sequence and adapter contamination is less than 0.1% in all samples.

Table 4. MultiQC quality metrics of sequencing data

MulitQC quality metrics in each sample type			
Sample	ORP	Protrudin	Control
Mean GC%	51	51	51
Mean per base N content	<5%	<5%	<5%
Mean adapter content	<0.1%	<0.1%	<0.1%
Mean sequence length	>70	>70	>70

The above metrics show that the sequencing data is of high quality and reliable, but to make the data even more reliable the sequences were subject to Trimmomatic quality control. Mainly all bases with a phred score lower than 30 were deleted, which removed roughly 0.6% of all bases.

Exploratory data analysis was performed before further analysis, to see whether or not there was any separation and grouping between the samples. First a principle component analysis (PCA) -plot was generated to visualise if there was any variation between the samples. As can be seen from figure 9, the samples are separating well and most samples are closely clustered together within their own group, except for the control samples, which show more separation within the group than the other sample types. Figure 9 also shows that the separation between the silenced samples is sufficient, therefore we can draw conclusions on the effects of silencing of either ORP2 or protrudin. Figure 9 also shows that protrudin samples differ more from the controls than the ORP2 samples.

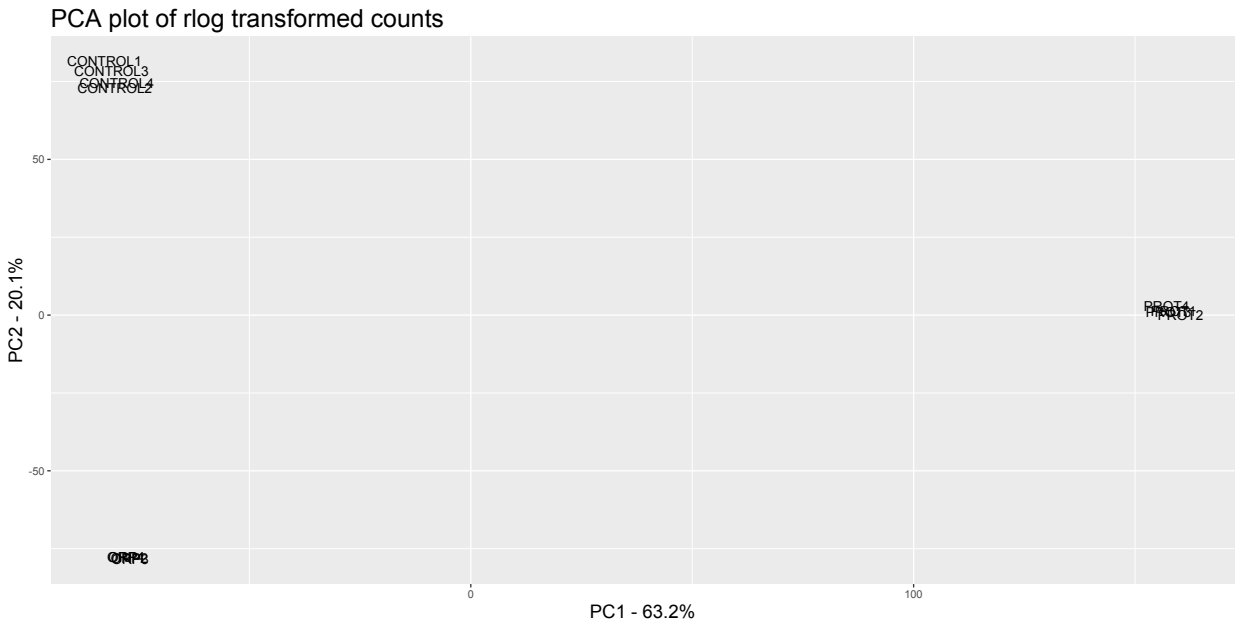


Figure 9. PCA plot of rlog normalized counts.

The same effect can be seen in the hierarchical clustering heatmap, which plots the variation in the 100 most variable genes in the samples (figure 10).

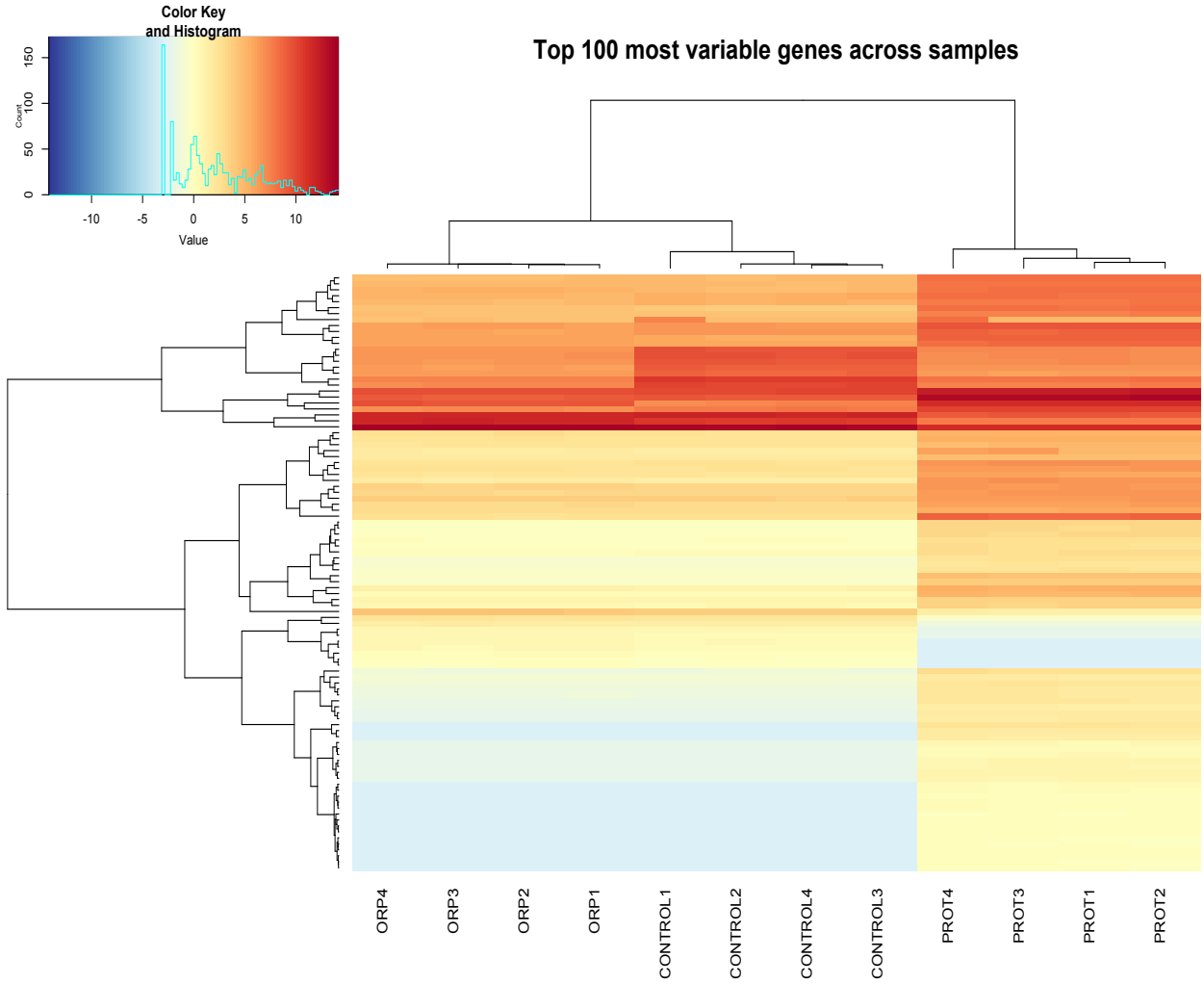


Figure 10. Heatmap of 100 most variable genes from rlog normalized counts in all samples.

Both the PCA plot and the heatmap show separation between the samples, especially the silenced samples from the controls. Grouping of similar samples is also distinct, therefore we can assume that we can differentiate samples from controls, and each other.

Next two volcano plots were produced to visualize the distribution of genes relative to the -log10 of the adjusted p-value and log2 -fold change, in the silenced samples. As we can see from figures 11 and 12, the fold change has not been substantial in most genes and we can see most of them cluster between the fold changes -1 – 1.

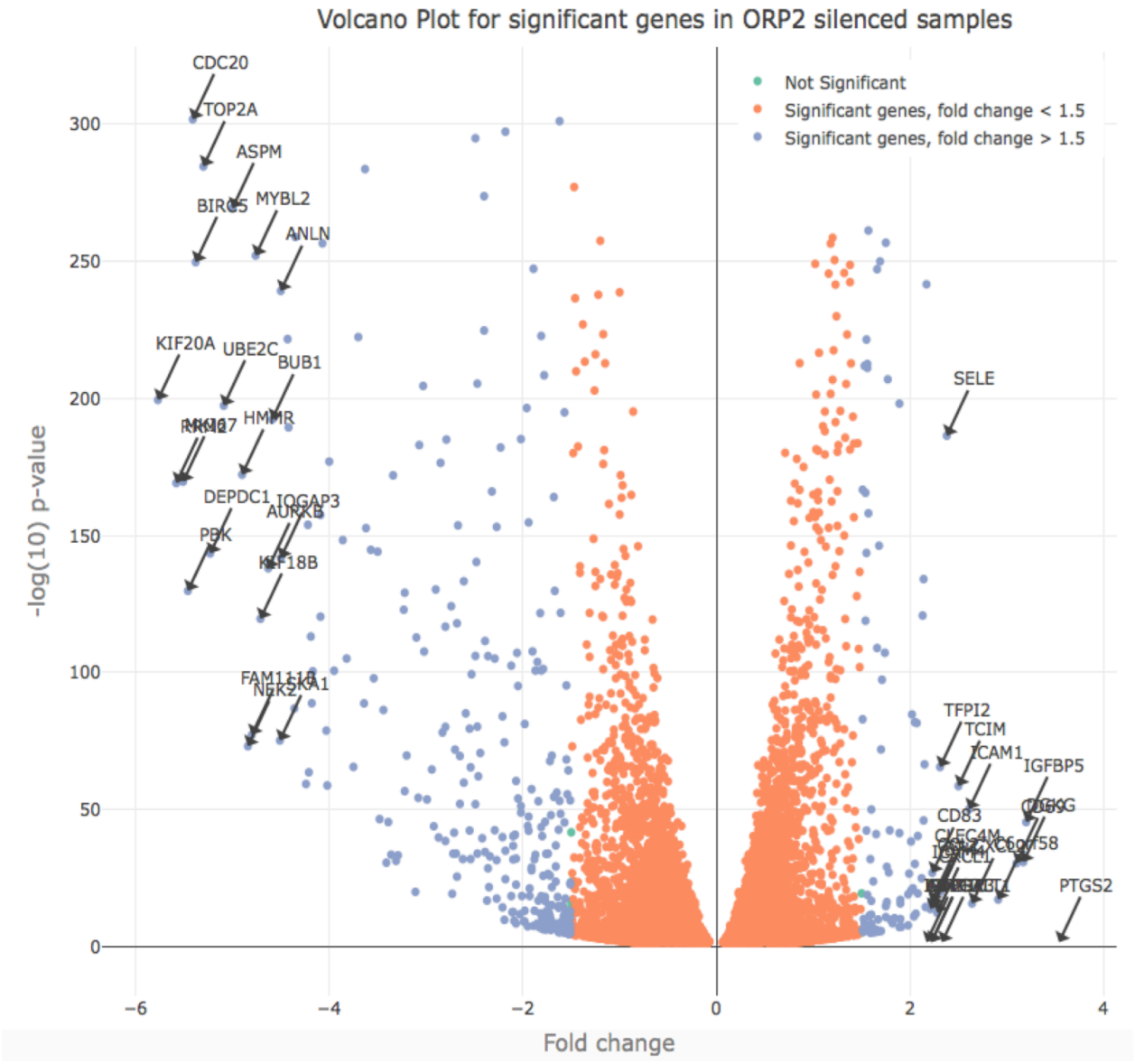


Figure 11. Volcano plot of significantly changed genes in ORP2 silenced samples, where x-axis represents the log2 fold change and y-axis depicts the -$\log_{10}$ of the adjusted p-value. Top 20 down- and upregulated genes have been annotated with arrows. Blue dots represent genes with absolute log2 fold change of more than 1.5, orange dots represent genes with absolute log2 fold change of less than 1.5

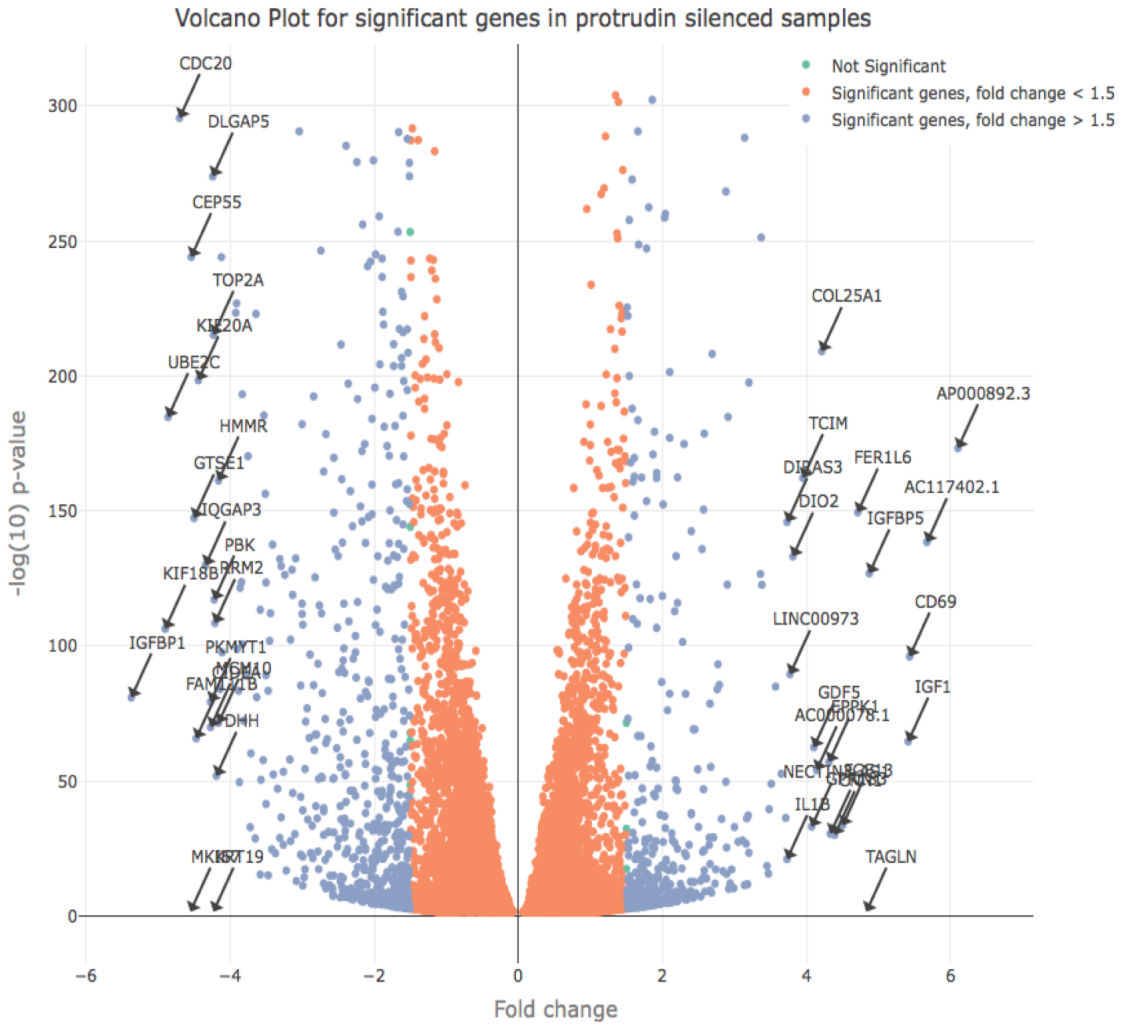


Figure 12. Volcano plot of significantly changed genes in ORP2 silenced samples where x-axis represents the log2 fold change and y-axis depicts the -$\log_{10}$ of the adjusted p-value. Top 20 down- and upregulated genes plotted and annotated with arrows. Blue dots represent genes that with absolute log2 fold change of more than 1.5, orange dots represent genes with absolute log2 fold change of less than 1.5

The fact that most genes cluster between -1 – 1 log fold change is due to assumptions in the DESeq2 algorithm, mainly the expectation that most genes are not differentially expressed, and the variation of low expressing genes is reduced with normalisation. The top 20 most affected genes by the KD of ORP2 or protrudin can be seen plotted in figures 11 and 12. These plots were generated from data where genes that were flagged as shRNA off-targets had been removed, see section “4.2 Prediction of shRNA off-target effects” for more information. There seems to be a difference in the distribution of genes in the volcano plots between the samples. In ORP2 more genes seem to have negative log2 fold changes than in protrudin samples where the distribution is more equal. There

also seems to be more downregulated genes with fold changes smaller than -1.5 in ORP2 samples than in protrudin samples. It is also notable that in ORP2 samples there are fewer genes with positive fold changes and the largest positive fold changes have relatively smaller adjusted p-values, whereas in protrudin samples there seems to be more variance in the adjusted p-values. There does not seem to be an obvious pattern in the top 20 most down- and upregulated genes in either of the sample types , but there seem to be a few kinases and genes related to cell cycle within them. Table 6 illustrates the expression and adjusted p-values of oxysterol binding proteins in both samples

Table 5. Expression of oxysterol binding proteins in ORP2 and protrudin KD samples. Differences in expression between the samples have been highlighted in yellow and genes upregulated in both samples have been highlighted in red

Expression of oxysterol binding proteins that were significantly differentially expressed in ORP2 and protrudin KD samples				
Symbol	Log2 fold change		Adjusted p-value	
Sample	ORP2	protrudin	ORP2	protrudin
OSBP	- 0,28	NA	1,86E-09	NA
OSBP2	- 0,39	- 1,3	1,17E-05	1,07E-39
OSBPL1A	- 0,74	- 0,44	4,74E-32	6,29E-13
OSBPL2	- 1,32	0,45	5,41E-89	4,21E-17
OSBPL3	- 0,5	- 0,92	2,43E-23	4,68E-67
OSBPL5	NA	- 0,52	NA	6,54E-15
OSBPL6	- 0,44	- 0,76	9,34E-05	2,46E-11
OSBPL7	0,42	NA	0,02091	NA
OSBPL8	0,65	1,08	8,70E-52	9,52E-125
OSBPL9	- 0,27	0,23	6,33E-11	2,05E-07
OSBPL10	0,51	0,27	2,16E-32	6,67E-09
OSBPL11	NA	- 0,46	NA	6,58E-14

There has been clear downregulation in many of the oxysterol binding proteins in both samples. The pattern of expression seems to be very similar in both samples and only *OSBPL2* and *OSBPL9* differ between them. Differences in expression of *OSBPL2* are most likely due to the KD of ORP2.

Prediction of shRNA off-target effects

Prediction of shRNA off-target effects was done with an inhouse developed R script, which can be found in appendix 1, or with GESS. List of top 10 genes in which the seed sequence was found can be found in table 7. There seems to be no abnormalities in the distribution of these adjusted p-values and log2 fold changes. There also appears to be no apparent pattern in the genes where the seed sequence was found.

Table 6. Ten most downregulated genes with the seed sequence in ORP2 and protrudin samples

Top 10 downregulated genes where the shRNA seed sequence was found in ORP2 and protruding KD samples					
ORP2			Protrudin		
Gene symbol	Log2 fold change	Adjusted p-value	Gene symbol	Log2 fold change	Adjusted p-value
GTSE1	-4,09	4,03E-158	*ELMOD1*	-3,72	1,57E-33
SRL	-1,98	9,72E-10	*CDC6*	-3,5	3,83E-124
SDCBP2	-1,66	2,57E-24	*ACOT11*	-2,84	4,01E-86
APOL4	-1,55	4,26E-09	*CLSPN*	-2,68	1,29E-73
USP2	-1,46	5,68E-07	*EMP2*	-2,42	9,43E-100
PLAC8	-1,41	8,63E-17	*GNAI2*	-2,06	<8,49E-260
CDH4	-1,33	6,62E-06	*DLAT*	-1,93	8,49E-260
SZRD1	-1,22	1,78E-238	*C1RL*	-1,76	1,04E-121
ELK3	-1,04	1,51E-98	*INSIG1*	-1,7	1,67E-160
RHPN2	-0,97	1,00E-05	*CYSLTR2*	-1,7	1,24E-05
Total number of significantly downregulated genes with the seed sequence		80	Total number of significantly downregulated genes with the seed sequence		96

GESS, a more robust shRNA off-target predictor, was also used to find possible off-targets. Results of top ten most downregulated unique genes found by the GESS algorithm can be found from table 8. The obvious difference between the inhouse developed R script and GESS, is that there are fewer off-targets found with the GESS algorithm. The less robust R scripts finds approximately 100 total off-targets from only the significantly downregulated genes, whereas GESS finds approximately 45 off-targets from every known transcript in the ENSEMBL database.

Table 7. Top 10 GESS identified downregulated off-targets in significantly changed genes.

GESS predicted, top 10 significantly downregulated, shRNA off-targets in ORP2 and protrudin silenced samples					
ORP			Protrudin		
Gene symbol	Log2 fold change	Adjusted p-value	Gene symbol	Log2 fold change	Adjusted p-value
KLF8	-2,16	3,63E-30	*KLF8*	-2,06	2,57E-26
USP2	-1,46	5,68E-07	*KCNN3*	-1,82	2,14E-35
KCNN3	-0,96	7,15E-15	*CD47*	-0,96	1,71E-72
ZFP36L1	-0,86	1,89E-97	*USP2*	-0,86	5,60E-03
LPP	-0,83	2,90E-32	*BCL10*	-0,65	5,71E-36
MEX3D	-0,82	4,20E-110	*ZNF704*	-0,63	7,44E-23
HDAC4	-0,66	1,95E-17	*SSR1*	-0,46	6,02E-29
BCL10	-0,56	2,23E-33	*UCHL5*	-0,38	9,89E-10
UCHL5	-0,54	1,64E-20	*ERN1*	-0,32	1,58E-07
HOOK3	-0,51	3,06E-13	*ZNHIT3*	-0,24	3,39E-04
Total number of off-targets in significantly downregulated genes		15	Total number of off-targets in significantly downregulated genes		13

These results show that there are only a few off-targets for each shRNA, and they do not seem to be affecting many critical biological pathways or functions, therefore further analysis can be assumed to be essentially unaffected by the off-target effects of shRNA.

Gene ontology overrepresentation analysis

Cellular proliferation seems to be affected in both samples. These terms seem to be heavily skewed towards cellular processes that are associated with the overall health of the cells. For example, in both protrudin and ORP2 samples, mRNAs involved in cell motility, cell cycle, cell division and mitotic cell cycle are significantly affected. These examples can be seen in table 9. GO analysis also shows a clear enrichment in terms related to angiogenesis, although the enrichment is not as large as compared to the aforementioned examples, but there are statistically significant changes in these GO terms.

Table 8. Biological function GO terms, their descriptions, FDR and enrichment of terms that are associated with cell cycle, protein phosphorylation and angiogenesis in ORP2 and protrudin KD samples.

Selection of relevant, significantly changed GO-terms in ORP2 and protrudin KD samples			
GO term	Description	FDR	Enrichment
GO terms in ORP2 KD samples			
GO:0000278	mitotic cell cycle	7,92E-07	4,47
GO:0008283	cell proliferation	5,32E-03	2,16
GO:0048870	cell motility	3,69E-04	1,69
GO:0007049	cell cycle	1,16E-03	2,28
GO:0051301	cell division	1,29E-10	4,32
GO:0006468	protein phosphorylation	1,75E-02	1,77
GO:0001525	angiogenesis	2,86E-04	2,15
GO:0045765	regulation of angiogenesis	8,32E-04	2,13
GO:1903670	regulation of sprouting angiogenesis	3,78E-02	2,77
GO:0045766	positive regulation of angiogenesis	3,97E-02	2,19
GO:1904018	positive regulation of vasculature development	1,95E-02	2,19
GO:1901342	regulation of vasculature development	3,76E-04	2,13
GO:0036293	response to decreased oxygen levels	3,56E-02	1,96
GO:0070482	response to oxygen levels	1,20E-02	1,95
Total number of enriched GO terms			275
GO terms in protrudin KD samples			
GO:0000278	mitotic cell cycle	6,06E-11	3,65
GO:0008283	cell proliferation	5,43E-07	1,89
GO:0048870	cell motility	2,35E-05	1,65
GO:0007049	cell cycle	1,43E-04	1,95
GO:0051301	cell division	6,02E-07	2,13
GO:0001932	regulation of protein phosphorylation	1,00E-08	2,11
GO:0001525	angiogenesis	1,02E-04	2,27
GO:0045765	regulation of angiogenesis	2,87E-02	2,04
GO:1903670	regulation of sprouting angiogenesis	3,11E-03	6,27
GO:0090049	regulation of cell migration involved in sprouting angiogenesis	5,18E-03	7,45
Total number of enriched GO terms			293

Other biological processes that seem to be similar between ORP2 and protrudin KD samples are related to the phosphorylation of proteins.

There seem to be significant changes in biological processes involved in cell cycle, angiogenesis and protein phosphorylation in both sample types. Therefore, there is cause to suspect that there might also be significant changes within the pathways that are related to these processes. Since GO-analysis does not predict whether or not the biological process has been up- or downregulated, gene set enrichment analysis has to be performed in order to find the molecular changes within these biological processes.

Gene set enrichment analysis with Wikipathways gene sets

Gene set enrichment analysis was performed to see which pathways have changed the most due to silencing. Table 10 shows the relative change of the pathways compared to control. These results show that pathways related to angiogenesis are significantly upregulated in both samples. Both samples show significant upregulation in the VEGFA – VEGFR2 signaling pathway, PI3K-Akt and MAPK signaling pathways. These results would suggest that there is significant upregulation in angiogenesis since these three pathways are intertwined and their upregulation is predicted to increase angiogenesis. The focal adhesion pathway also shows significant upregulation in both samples, which could be argued to be related to angiogenesis. Samples differ in the upregulation of their pathways as in ORP2 the angiogenesis and matrix metalloproteinases pathways are significantly upregulated, whereas in protrudin samples the brain-derived neurotrophic factor (BDNF) signaling and PDGFB pathways are upregulated. These results are in line with the results acquired from gene ontology overrepresentation analysis.

Most notable downregulated pathway in both samples is the cholesterol biosynthesis pathway and pathways that are related to cell cycle, such as cell cycle pathway, DNA replication, G1 to S cell cycle control pathways, to name a few. Pathways related to cell cycle seem to be consistent with the results obtained from the gene ontology overrepresentation analysis. There seems to be a difference between the samples in some metabolic pathways, mainly that the protrudin samples have significant downregulation in glycolysis related pathways and other lipid metabolism pathways, whereas ORP samples seem to have more downregulation in amino acid related pathways. Good examples of these are the TCA cycle, glycolysis and gluconeogenesis pathways in protrudin samples, and amino acid metabolism, urea cycle and metabolism of amino groups pathways in ORP samples. A selection of these pathways and their p- and q-values can be found in table 10. Appendix 3 figures 1 and 2 illustrate all up- and downregulated pathways in each sample, which have a p-value of less than 0.01.

Table 9. Selection of five down- and upregulated gene sets in ORP2 and protrudin silenced samples.

Selection of top 5 up- and downregulated pathways in ORP2 and protrudin KD samples			
Pathway	p-value	q-value	regulation
Pathways affected by the ORP2 silencing			
Cell Cycle	1,49E-21	2,46E-19	-
DNA Replication	5,57E-17	4,60E-15	-
Cholesterol Biosynthesis Pathway	6,68E-08	2,45E-06	-
G1 to S cell cycle control	1.24E-12	8.20E-11	-
Amino Acid metabolism	9,26E-07	2,67E-05	-
VEGFA-VEGFR2 Signaling Pathway	1,74E-07	1,91E-05	+
Focal Adhesion-PI3K-Akt-mTOR-signaling pathway	6,11E-06	2,01E-4	+
MAPK Signaling Pathway	2,25E-05	5,70E-4	+
Matrix Metalloproteinases	1,60E-3	1,20E-2	+
Angiogenesis	7,17E-3	3,64E-2	+
Pathways affected by the protrudin silencing			
DNA Replication	5,75E-16	9,72E-14	-
Cell Cycle	1,61E-14	1,36E-12	-
G1 to S cell cycle control	6,96E-7	2,13E-5	-
Cholesterol Biosynthesis Pathway	4,67E-4	5,44E-3	-
Glycolysis and Gluconeogenesis	2,31E-5	4,59E-4	-
Brain-Derived Neurotrophic Factor (BDNF) signaling pathway	4,24E-5	1,60E-3	+
VEGFA-VEGFR2 Signaling Pathway	4,00E-4	7,12E-3	+
Focal Adhesion-PI3K-Akt-mTOR-signaling pathway	8,63E-4	1,12E-2	+
MAPK Signaling Pathway	1,25E-3	1,30E-2	+
PDGFR-beta pathway	7,59E-3	4,20E-2	+

Since the most interesting pathways related to the hypothesis of this thesis are the angiogenesis, PI3K-Akt, MAPK and cholesterol biosynthesis pathways, the gene expression of the genes within these pathways was plotted onto pathway maps with PathVisio. Illustrated in figure 13 is the simplified angiogenesis pathway map, which also includes parts of the PI3K-Akt and MAPK pathways.

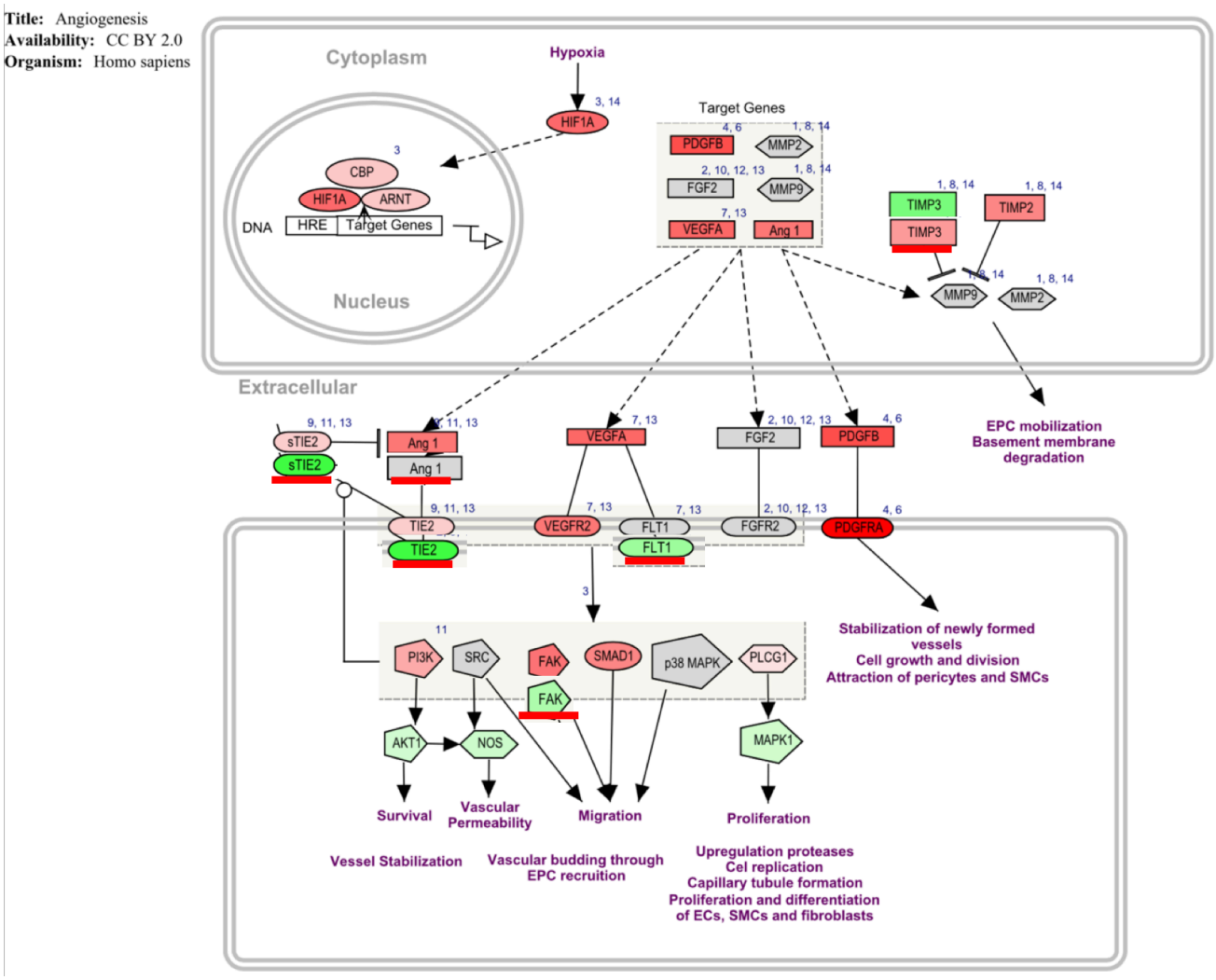


Figure 13. Simplified pathway map of the main components related to angiogenesis. Differences in the expression of genes between ORP2 and protrudin samples have been displayed in the following manner: Top most gene symbol represent the gene expression in ORP2 samples and below is displayed its counterpart in protrudin samples, which has also been underlined with red. For example, the sTIE2 and TIE2 genes have been upregulated in ORP2 samples but downregulated in protrudin samples. Significantly changed genes, have been plotted on the map and their relative fold changed is displayed as red for upregulation and as green for relative downregulation

The angiogenesis pathway map shows that the *Akt1*, *NOS* and *MAPK1* genes are downregulated in both samples, which could suggest reduced angiogenesis. They also show that most upstream genes such as *PI3K*, *SMAD1* and *PLCG1* are upregulated in both samples. The angiogenesis pathway differs in these two samples in the regulation of *FAK*, *TIE2*, *FLT1* and *Ang1*.

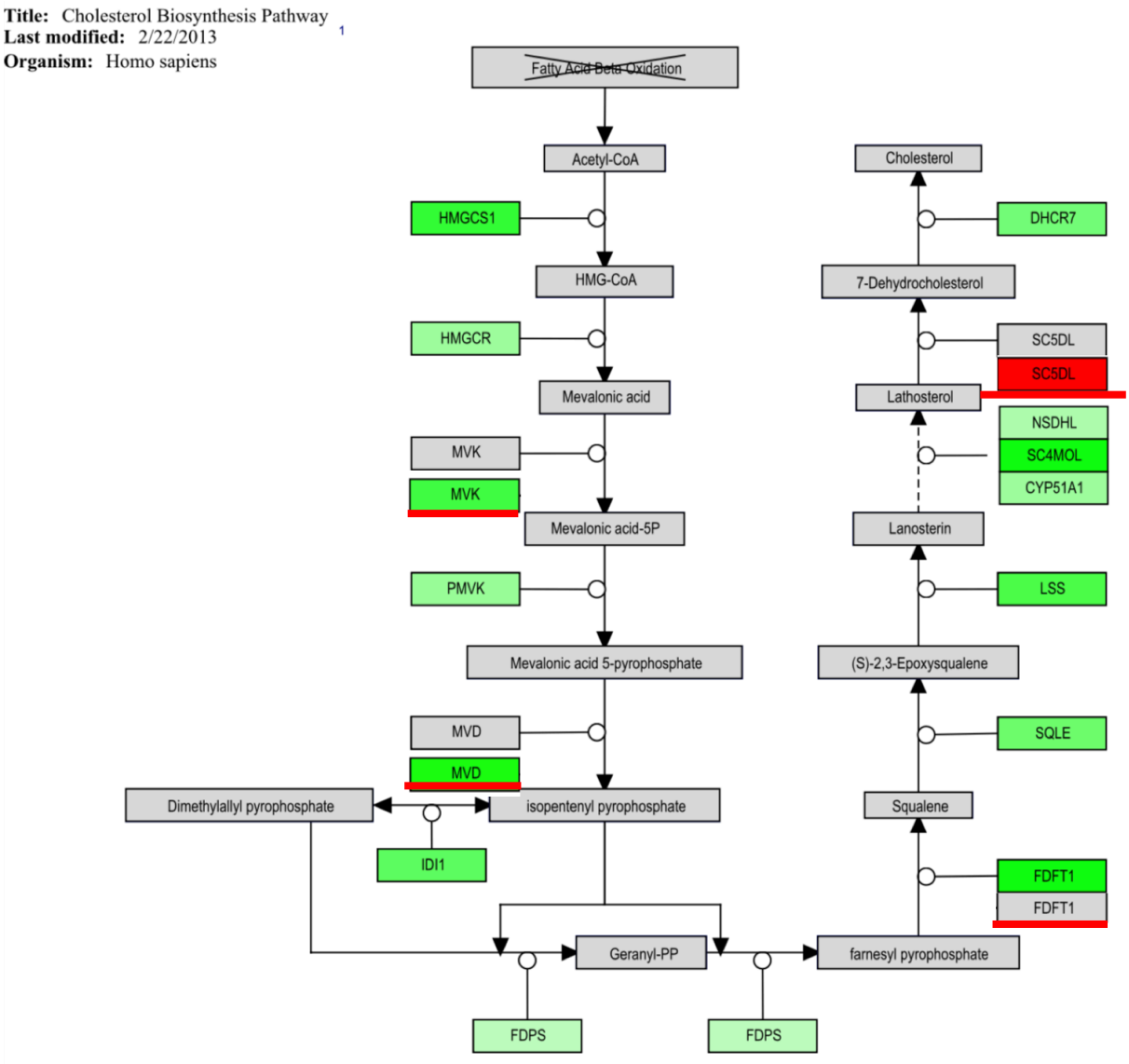


Figure 14. Cholesterol biosynthesis pathway map. Differences in the expression of genes between ORP2 and protrudin samples have been displayed in the following manner: Top most gene symbol represent the gene expression in ORP2 samples and below is displayed its counterpart in protrudin samples, which has also been underlined with red. For example, the SC5dl has not been differentially expressed in ORP2 samples but it is upregulated in protrudin samples. Significantly changed genes have been plotted on the map and their relative fold change is displayed as red for upregulation and as green for relative down regulation.

The cholesterol biosynthesis pathway in figure 14, shows clear downregulation in both samples and only one gene, *SC5DL,* in protrudin samples is upregulated.

6 Discussion

Only the major results of this thesis will be discussed in the same order as they appear in the results section. This section will be presented as guiding headers that are either similar or identical to the sub headers of the results section.

Exploratory data analysis

Exploratory data analysis provided results that could be expected from a standard RNA-sequencing experiment. Considering that there are thousands of genes that are used in PCA and heatmaps to differentiate the samples, there is little chance that there would be no clear separation between the samples. This is due to the enormous volume of data that is used for the differentiation. As mentioned already, the variation in the top 100 most variable genes seems to be very similar within all samples, but this is most likely due to normalisation that reduces variation in the gene expression data. This effect can be seen the DESeq2 dispersion estimation in figure 15, which shows the real distribution of genes in black, normalized distribution in blue and the fitted curve, which is used to produce the normalisation.

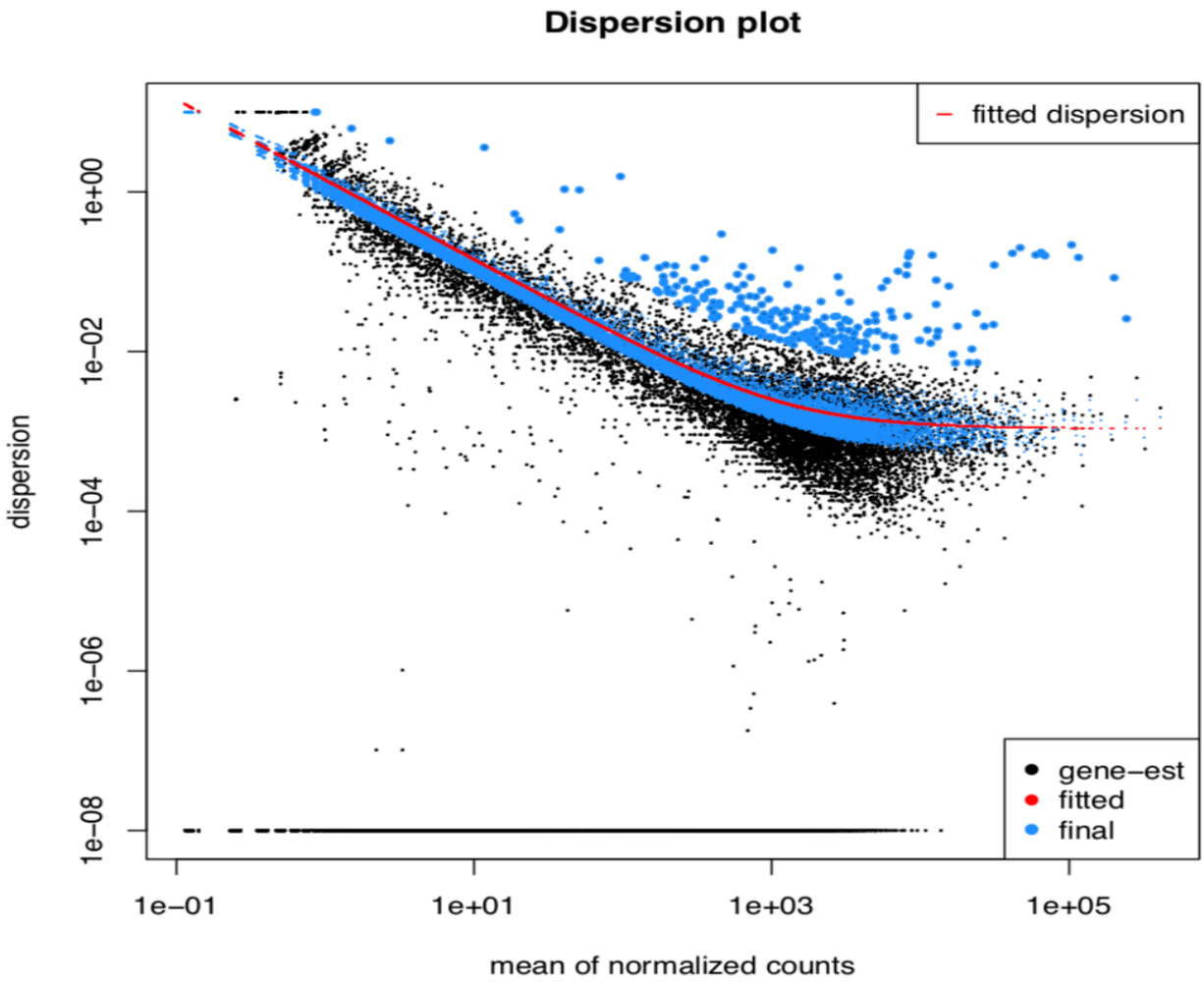


Figure 15. Dispersion estimation of ORP2 samples that shows the original distribution of gene expression in black, normalized distribution in blue and the fitted curve used to normalize the expression data.

The overall distribution of the gene expression results is as expected. The expression changes for most genes are not drastic and as noted before, most of the significantly changed genes cluster between fold changes of -1 – 1. We can see that most transcripts have not been significantly changed as only around 5% (~10 000) of all known transcripts (208 689) have changed significantly. This result could also be due to the bias of DESeq2, which assumes that most genes are not differentially expressed. Why the expression of genes in ORP2 silenced samples, displayed in the volcano plot, is more skewed toward downregulation and why the most upregulated genes have such similar smaller $-\log_{10}$ adjusted p-values, could be just due to chance. This effect could also be due to the silencing of ORP2, which could affect the cells transcriptional machinery and therefore there would be less transcripts in general. The same effects seen in the distribution of

genes in ORP2 cannot be seen in protrudin samples, which is very standard and there are no obvious patterns in the distribution of differentially expressed genes.

Examination of the regulation of oxysterol binding proteins in both samples are in a stark contrast to a previous study performed by Kentala *et al.* (Kentala, Koponen, Kivelä, et al., 2018) who showed that the knock-out of ORP2 did not affect the expression of other oxysterol binding proteins in a significant way, whereas these results show that there has been significant downregulation of many oxysterol binding proteins in both samples. The downregulation of other oxysterol binding proteins could be a problem for downstream analysis since it could be hard to examine the effects of ORP2 KD if other similar proteins are also affected.

Prediction of shRNA off-target effects

Off-target prediction of shRNA is not a straightforward task and number of different variables have to be considered, in order to make confident predictions on the off-target effects. This fact is evident when the two different algorithms are compared, and this comparison clearly shows that probing for putative miRNA-like seed sequence targets in the 3'UTR regions of transcripts is not robust enough method to produce accurate predictions. The GESS algorithm is able to take into account more variables and it is able to perform statistical analysis in the prediction, therefore it identifies less off-targets. Controlling for off-target effects by removing any downregulated genes identified by GESS or the inhouse R script, might not be enough in some situations. There could be profound effects in biological functions if one of the off-targets was a gene that is crucial for a pathway, for example, unintentional KD of VEGFR2, which would affect the VEGFA-VEGFR2 signaling pathway in unpredictable ways. Out of the off-targets predicted by GESS, genes *LPP* and *CD47* are of most concern. Lipoma preferred partner (*LPP*) is associated with the promotion of motility and permeability in ECs (Leung et al., 2018) and Cluster of Differentiation 47 (*CD47*) is associated in apoptosis, proliferation, adhesion and migration as well as angiogenesis (Sick et al., 2012). *CD47* downregulation can also lead to increased VEGFR2 downstream signaling (Xing et al., 2009). *CD47* downregulation can only be seen in protrudin samples, and it is slightly upregulated in ORP2 samples. If indeed *LPP* has been downregulated due to off-target effects of the shRNA, there should

be downregulation in motility but instead mRNAs related to migration and motility seem to be upregulated. In the case of *CD47* downregulation angiogenesis should be induced due to the reduction of thrombospondin-1 (THBS1) CD47 complexes, which inhibit cell migration, tube formation as well as NO stimulated response in ECs (Sick et al., 2012). The log2 fold change of *THBS1* has been 0.63 and 0.61 in ORP2 and protrudin samples, respectively. The expression of *THBS1* seems to be very similar in both samples and one could expect some compensatory effects in the expression of *THBS1* in protrudin samples, if *CD47* were abnormally downregulated. Previous experiments with knock-out ORP2 Huh7 cells by Kentala *et al.* (Kentala, Koponen, Kivelä, et al., 2018; Kentala, Koponen, Vihinen, et al., 2018) have also provided evidence that the knock-out of ORP2 reduces cellular proliferation, migration and cell adhesion. A final conclusion on whether or not *LPP* and *CD47* have been downregulated due to off-target effects cannot be made based on this data. However, a conclusion that their downregulation has likely not had a significant effect on the cellular functions of interest can be made; mostly due to the evidence provided by Kentala *et al.* and the fact that *THBS1* expression is the same in both samples, even though *CD47* has been downregulated only in protrudin KD samples.

Gene ontology overrepresentation analysis

GO-analysis suggests clear perturbation in the health of cells. This is very obvious from the enriched GO-terms, which all show a clear connection to biological functions that are related to the health of cells. These effects cannot be attributed to the lentiviral transduction nor puromycin selection of cells, since control cells were subject to the same conditions. Therefore, there should not be any enrichment of these GO-terms due to the experimental protocol, and we have to assume that these results are due to the silencing. Since GO-terms that are related to angiogenesis in both samples and response to oxygen levels in ORP2 samples are overrepresented, these results would indicate that there have been major changes in genes related to angiogenesis. At first glance the GO-terms in ORP2 samples would seem to indicate that angiogenesis has been increased, since terms such as positive regulation of angiogenesis and response to decreased oxygen levels (i.e. hypoxia) are overrepresented. However, conclusions on the direction of regulation of the genes cannot be made based on overrepresentation of GO-terms alone and a more detailed analysis on the gene expression is required. The results would also

seem to indicate that there has been no perturbations in the transport of lipids, since there is no overrepresentation of GO-terms that are related to the transport and synthesis of lipids. Furthermore, the overrepresented GO-terms for ORP2 are similar to the ones obtained by Kentala *et al.* (Kentala, Koponen, Kivelä, et al., 2018). Most of the differences between the data presented here and data obtained in the aforementioned study are most likely due to the analysis being performed with Ingenuity Pathway Analysis (QIAGEN), since the curated gene sets might be different and also have different names. A more detailed analysis in to the genes and pathways related to the enriched GO-terms has to be made in order to find possible molecular changes.

Gene set enrichment analysis with Wikipathways gene sets

Examination of the GAGE predicted up- and downregulated pathways seem to indicate that angiogenesis has been upregulated. For example, angiogenesis related pathways such as the matrix metalloproteinases, VEGFA-VEGFR2 signaling pathway and focal adhesion-PI3K-Akt-mTOR-signaling pathway are all upregulated. There also seems to be some evidence to support the fact that angiogenesis has been perturbed in the downregulated pathways. For example, pathways related to the proliferation of cells seem to be downregulated, which would indicate that even though the cells are prompted to proliferate through the angiogenesis related pathways, they are not able to do so. Also, one of the indicators of angiogenesis is the switch from aerobic cell respiration to glycolysis in ECs, but the glycolysis pathway seems to be downregulated instead of upregulation. There also seems to be an effect in the synthesis of lipids, which was not seen in the GO-analysis, since the cholesterol biosynthesis pathway has been downregulated. This could be due to the lack of transport of cholesterol from the ER, which would increase the amount of cholesterol in the ER, and therefore the transcription of cholesterol biosynthesis genes is downregulated. There could also be some perturbation in the sensing of cholesterol at the ER and therefore the pathway has been downregulated. Also, the downregulation of *SREBF2* and *SREBF1* (transcription factors affecting the transcription of enzymes that synthesise sterols), which have fold changes of -0.5 or 0.09 and -0.2 or -0.77 in ORP2 and protrudin samples, respectively, could explain the reduction of cholesterol synthesis.

When the simplified angiogenesis pathway map is examined more closely, we can see that in fact the effect in this pathway has been the downregulation of downstream elements such as *Akt1*, *NOS* and *MAPK1*, which would indicate that angiogenesis has been reduced. Even though GAGE has flagged this pathway as upregulated, which is most likely due to the upregulation of most upstream elements in the pathway, such as *VEGFA, HIF1A, VEGFR2, PDGFRA, PI3K* and *Ang 1*, the aforementioned downstream elements are downregulated, and therefore the actual effect is likely reduced angiogenesis. This effect could be due to compensatory mechanisms in the cell, which are trying to initiate angiogenesis through these upstream elements. This compensatory effect is probably due to the fact that proliferation has been reduced along with the expression of the VEGFR1, and therefore VEGF stimulation is not being resolved. Since the VEGF stimulation is not being reduced by stalk cells producing VEGFR1, most cells are prompted to enter a tip cell-like phenotype, which could also explain the upregulation of *VEGFR2* and *PDGFRA*. ECs are also prompted to migrate as either *FAK, SMAD1* or *p38 MAPK* have been upregulated in both samples, even though proliferation has been reduced. This effect could also be explained by ECs transforming to the aforementioned tip cell-like phenotype, which would prompt the upregulation of migration related genes as well as the focal adhesion pathway. The downregulation of *Akt1* and *MAPK1* is the most likely explanation for the downregulation of the cell cycle, DNA replication and G1 to S cell cycle control pathways. This effect can be explained by the downregulation of encoding cyclin D1 (*CCDN1)*, among other cyclins, which is part of the PI3K-Akt pathway, and *Akt1* is one of the downstream elements of this pathway. *MAPK1*, along with other downregulated MAPKs, such as *MAPK8* and 9, also control the cell cycle through *CCDN* genes. Since the cells have are not entering S phase from G1 phase due to the downregulation of CCDN genes, it is quite reasonable to suggest that the DNA replication pathway is downregulated due to this effect. The downregulation of *Akt1* and *MAPK1* could be explained with the ECs changing into the tip cell-like phenotype, which would prompt the cells reduce proliferation and therefore reduce signaling through the PI3K-Akt and MAPK pathways.

As an alternative, the effects seen in the aforementioned pathways could also be due to the downregulation of other oxysterol binding proteins and the lack of endosomal transport to the PM due to reduced expression of protrudin. Among the downregulated

ORPs there are at least four (OSBPL3, OSBPL5, OSBPL6, OSBPL7), which bind cholesterol and/or PtdInPs as well as localise to the ER and PM and two ORPs, which bind to similar lipids and localise to endosomes as well as other membranes (OSBPL1A, OSBPL11*)* (Pietrangelo & Ridgway, 2018). This could result in the lack of precursors or signaling lipids at the PM because their transport has been reduced due to the lack of lipid transport between ER and endosomes, as well as the reduced endosomal transport to the PM. This could in turn result in the compensatory effect of the upregulation of signal mediators, such as PI3K, which would try to compensate for the lack of lipids by amplifying the speed of signal production.

Another alternative hypothesis for the downregulation of the PI3K-Akt and MAPK pathways in ORP2 samples is related to the previously studied effects of ORP2 interactions with Akt1 (Kentala, Koponen, Kivelä, et al., 2018; Kentala, Koponen, Vihinen, et al., 2018). Kentala *et al.* showed that ORP2 has a critical role in the formation of a complex that phosphorylates Akt1, which in turn affects multitude of cellular functions, such as proliferation, lamellipodia formation, migration, adhesion as well as glucose uptake and glycolysis. Therefore, the upregulation of the upstream elements in the PI3K-Akt pathway could be compensatory effects due to the reduced downstream signaling through Akt1.

Conclusions

This thesis has provided evidence that the KD of ORP2 and protrudin have had significant and profound changes in mRNAs related to cholesterol biosynthesis, cell cycle, cell division and angiogenesis. These changes in mRNA levels, have most likely had a reducing effect in these biological functions. A common factor between the sample types that could explain these effects is the transformation of ECs to a tip cell-like phenotype, which would reduce proliferation and increase migration in ECs. In ORP2 silenced samples these effects could be due to the reduced phosphorylation of Akt1 and in protrudin samples these effects might be attributed to the reduced amount of endosomal trafficking from the ER to the PM. Results obtained in this thesis also indicate that ORP2 KD has similar effects at the mRNA level in HUVECs as ORP2 knock-out has in Huh7 cells (Kentala, Koponen, Kivelä, et al., 2018; Kentala, Koponen, Vihinen, et al., 2018).

Future prospects

Verifying gene expression could be the first step towards further experimentation. RT-qPCR could be used to examine the expression levels as an independent verification of the RNA-sequencing results. RT-qPCR does have some limitations and it is not able to detect as small changes as RNA-sequencing, but it is a very robust method to detect larger fold changes. RT-qPCR is also more error prone when compared to RNA-sequencing and considering that most sequencing technologies are well established and subject to rigorous quality control, RT-qPCR may not be well motivated. Changes in gene expression might not translate to changes in protein levels, therefore it is crucial to verify the protein level changes of key nodes within the discussed pathways. Verifying protein levels is simple enough with Western blotting, but complications might arise if these proteins have long half-lives and changes in protein levels might not be seen in the span of a typical 72-hour KD experiment. Most crucial protein levels to verify are those of Akt1, MAPK1 and NOS. If these proteins have indeed been reduced in the silenced samples, further experiments might be warranted.

There are at least three different assay types that can be used to study angiogenesis *in vitro*. These assays measure the rate of migration and proliferation as well as the differentiation of ECs. Proliferation assays are quite simple to perform, since determining net cell numbers can be done using simple equipment. These experiments are quite straightforward, and they involve growing both control and silenced ECs in media supplemented with growth factors such as VEGF, and each cell type is counted periodically to determine their doubling rate. Other methods to study cell proliferation include methods, such as the xCELLigence system, which measures the impedance between electrodes that increases as cells divide on a plate populated with microelectrodes. Studying cell-cycle kinetics or measuring the amount of new double stranded DNA formation with either radioactively labelled thymine or with bromodeoxyuridine (BrdU) ELISA kit, which measures the amount of BrdU incorporated to new DNA, could also be used (Tahergorabi & Khazaei, 2012).

Migration assays are more complex and involve more sophisticated equipment compared to a standard doubling rate proliferation assay. Most often a scratch assay or wound healing assay is used to quantify the migration rate of cells. This experiment is performed

by scoring a line in an EC monolayer, and cells around the scratch are monitored and their rate of migration to the empty area is measured. This measuring can be done periodically manually, or automated systems that track each cell can be employed (Tahergorabi & Khazaei, 2012).

More complicated functional assays are the differentiation assays. These assays are used to determine the ECs ability for sprouting or tubulogenesis. These assays can be performed on thin 2D like matrixes or in 3D matrixes such as GrowDex or Matrigel. The ultimate goal is to quantify the length and number of new sprouts or microvessle-resembling tubules growing within the matrix with a microscope (Tahergorabi & Khazaei, 2012). Recent unpublished findings by the Olkkonen group from Minerva Foundation Institute for Medical Research have shown that the KD of ORP2 or protrudin inhibit tubular network formation in HUVECs (Koponen, et al*.,* unpublished).

Confirming whether or not the cholesterol biosynthesis pathway is downregulated due to cholesterol being trapped at the ER could be studied with at least two different methods. One of these methods includes studying the cleavage and localisation of *SERBF2* (for example with Abnova, catalog number KA1378). Based on the localisation of *SERBF2* to either the cytosol or to the nucleus, conclusions on the transcription of cholesterol biosynthesis related enzymes could be made. Cholesterol position in the ER could be confirmed with a cholesterol efflux assay, which uses radiolabelled cholesterol and measures cholesterol movement from cells to extracellular cholesterol acceptors (Low, Hoang, & Sviridov, 2012).

7 Acknowledgements

I would like to thank Vesa Olkkonen for excellent guidance during the master's thesis work. His ideas were enlightening during the examination of the results and his input on the drafts of this thesis were greatly appreciated. The input of Amita Arora and Annika Koponen were also instrumental during the technical parts of this thesis and without their expertise this thesis would not have turned out this way. The input of Amita and Annika also helped me to interpret the results in ways that I would otherwise have missed. I would also like to acknowledge the University of Helsinki Functional Genomics sequencing unit

for providing excellent sequencing data and especially Juho Väänänen, whose expert input in the processing of sequencing data was extremely helpful. I also extend my gratitude to the Sigurd Juselius foundation that has been supporting this thesis with a generous grant. The contributions of my family and friends for the completion of thesis cannot be understated. Their kindness and support has been a driving force during my studies, and I am very grateful to have such amazing people to support me. Finally, I would also like to thank Rosanna Kuivalainen whose constant motivation and uplifting spirit were crucial for the completion of this thesis.

References

Alberts, B., Johnson, A., Lewis, J., Morgan, D., & Raff, M. (2014). *Molecular Biology of the Cell*. Retrieved from http://ebookcentral.proquest.com/lib/helsinki-ebooks/detail.action?docID=5320520

Alpy, F., Rousseau, A., Schwab, Y., Legueux, F., Stoll, I., Wendling, C., … Tomasetto, C. (2013). STARD3 or STARD3NL and VAP form a novel molecular tether between late endosomes and the ER. *Journal of Cell Science*, *126*(23), 5500–5512. https://doi.org/10.1242/jcs.139295

Bohler, A., Smeets, B., Willighagen, E. L., Cirillo, E., Mélius, J., Nunes, N., … Sinha, S. R. (2015). WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic Acids Research*, *44*(D1), D488–D494. https://doi.org/10.1093/nar/gkv1024

Bolger, A. M., Usadel, B., & Lohse, M. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brazma, A., Kasprzyk, A., De Moor, B., Davis, S., Durinck, S., Huber, W., & Moreau, Y. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, *21*(16), 3439–3440. https://doi.org/10.1093/bioinformatics/bti525

Brouwer, C., Blanchard Jr, S. G., Luo, W., Pant, G., & Bhavnasi, Y. K. (2017). Pathview Web: user friendly pathway visualization and data integration. *Nucleic Acids Research*, *45*(W1), W501–W508. https://doi.org/10.1093/nar/gkx372

Brouwer, C., & Luo, W. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, *29*(14), 1830–1831. https://doi.org/10.1093/bioinformatics/btt285

Chen, Y. A., Scheller, R. H., & Medical, H. H. (2001). SNARE-mediated membrane fusion. *Nature*, *2*, 98–106.

Chiapparino, A., Maeda, K., Turei, D., Saez-Rodriguez, J., & Gavin, A.-C. (2016). The orchestra of lipid-transfer proteins at the crossroads between metabolism and signaling. *Progress in Lipid Research*, *61*, 30–39. https://doi.org/10.1016/J.PLIPRES.2015.10.004

Copeland, D. E. (1959). An Association between Mitochondria and the Endoplasmic Reticulum in Cells of the Pseudobranch Gland of a Teleost. *The Journal of Cell Biology*, *5*(3), 393–396. https://doi.org/10.1083/jcb.5.3.393

Di Paolo, G., & De Camilli, P. (2006). Phosphoinositides in cell regulation and

membrane dynamics. *Nature*, *443*(7112), 651–657. https://doi.org/10.1038/nature05185

Dobin, A., Davis, C. A., Zaleski, C., Schlesinger, F., Drenkow, J., Chaisson, M., … Gingeras, T. R. (2012). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Donaldson, J. G., & Honda, A. (2005). Localization and function of Arf family GTPases. *Biochemical Society Transactions*, *33*(4), 639–642.

Drin, G. (2014). Topological Regulation of Lipid Balance in Cells. *Annual Review of Biochemistry*, *83*(1), 51–77. https://doi.org/10.1146/annurev-biochem-060713-035307

Duarte, A., Suchting, S., Freitas, C., Noble, F., Benedito, R., Bre, C., & Eichmann, A. (2007). The Notch ligand Delta-like 4 negatively regulates endothelial tip cell formation and vessel branching. *Proceedings of the National Academy of Sciences*, *104*(9), 3225–3230.

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, *4*(8), 1184–1191. https://doi.org/10.1038/nprot.2009.97

Eden, E., Lipson, D., Yogev, S., & Yakhini, Z. (2007). Discovering Motifs in Ranked Lists of DNA Sequences. *PLOS Computational Biology*, *3*(3), e39. Retrieved from https://doi.org/10.1371/journal.pcbi.0030039

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, *10*(1), 48. https://doi.org/10.1186/1471-2105-10-48

Geudens, I., & Gerhardt, H. (2011). Coordinating cell behaviour during blood vessel formation. *Development*, *138*(21), 4569–4583. https://doi.org/10.1242/dev.062323

Hammond, G. R. V, Fischer, M. J., Anderson, K. E., Holdich, J., Koteci, A., Balla, T., & Irvine, R. F. (2012). PI4P and PI(4,5)P2 Are Essential But Independent Lipid Determinants of Membrane Identity. *Science*, *337*(6095), 727–731.

Hellström, M., Phng, L., Hofmann, J. J., Wallgard, E., Coultas, L., Lindblom, P., … Betsholtz, C. (2007). Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis. *Nature*, *445*(7129), 776–780. https://doi.org/10.1038/nature05571

Hölttä-Vuori, M., Alpy, F., Tanhuanpää, K., Jokitalo, E., Mutka, A.-L., & Ikonen, E. (2005). MLN64 Is Involved in Actin-mediated Dynamics of Late Endocytic

Organelles. *Molecular Biology of the Cell*, *16*(8), 3873–3886. https://doi.org/10.1091/mbc.e04-12-1105

Jean, S., & Kiger, A. A. (2012). Coordination between RAB GTPase and phosphoinositide regulation and functions. *Nature Reviews Molecular Cell Biology*, *13*(7), 463–470. https://doi.org/10.1038/nrm3379

Kallio, M. A., Tuimala, J. T., Hupponen, T., Klemelä, P., Gentile, M., Scheinin, I., … Korpelainen, E. I. (2011). Chipster: user-friendly analysis software for microarray and other high-throughput data. *BMC Genomics*, *12*, 507. https://doi.org/10.1186/1471-2164-12-507

Kamburov, A., Pentchev, K., Galicka, H., Wierling, C., Lehrach, H., & Herwig, R. (2011). ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Research*, *39*(Database issue), D712–D717. https://doi.org/10.1093/nar/gkq1156

Kentala, H., Koponen, A., Kivelä, A. M., Andrews, R., Li, C., Zhou, Y., & Olkkonen, V. M. (2018). Analysis of ORP2-knockout hepatocytes uncovers a novel function in actin cytoskeletal regulation. *The FASEB Journal*, *32*(3), 1281–1295. https://doi.org/10.1096/fj.201700604R

Kentala, H., Koponen, A., Vihinen, H., Pirhonen, J., Liebisch, G., Pataj, Z., … Olkkonen, V. M. (2018). OSBP-related protein-2 (ORP2): a novel Akt effector that controls cellular energy metabolism. *Cellular and Molecular Life Sciences*, *75*(21), 4041–4057. https://doi.org/10.1007/s00018-018-2850-8

Koponen, A., Arora, A., Takahashi, K., Kentala, H., Kivelä, A. M., Jääskeläinen, E., … Olkkonen, V. M. (2019). ORP2 interacts with phosphoinositides and controls the subcellular distribution of cholesterol. *Biochimie*, *158*, 90–101. https://doi.org/10.1016/j.biochi.2018.12.013

Kutateladze, T. G. (2010). Translation of the phosphoinositide code by PI effectors. *Nature Chemical Biology*, *6*(7), 507–513. https://doi.org/10.1038/nchembio.390

Kutmon, M., van Iersel, M. P., Bohler, A., Kelder, T., Nunes, N., Pico, A. R., & Evelo, C. T. (2015). PathVisio 3: An Extendable Pathway Analysis Toolbox. *PLOS Computational Biology*, *11*(2), e1004085. Retrieved from https://doi.org/10.1371/journal.pcbi.1004085

Lemmon, M. A. (2008). Membrane recognition by phospholipid-binding domains. *Nature Reviews Molecular Cell Biology*, *9*(2), 99–111. https://doi.org/10.1038/nrm2328

Leung, C. S., Yeung, T.-L., Yip, K.-P., Wong, K.-K., Ho, S. Y., Mangala, L. S., … Mok, S. C. (2018). Cancer-associated fibroblasts regulate endothelial adhesion protein LPP to promote ovarian cancer chemoresistance. *The Journal of Clinical*

Investigation, *128*(2), 589–606. https://doi.org/10.1172/JCI95200

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Low, H., Hoang, A., & Sviridov, D. (2012). Cholesterol efflux assay. *Journal of Visualized Experiments : JoVE*, (61), e3810–e3810. https://doi.org/10.3791/3810

Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D., & Woolf, P. J. (2009). GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*, *10*(1), 161. https://doi.org/10.1186/1471-2105-10-161

Mesmin, B., Bigay, J., Moser von Filseck, J., Lacas-Gervais, S., Drin, G., & Antonny, B. (2013). A Four-Step Cycle Driven by PI(4)P Hydrolysis Directs Sterol/PI(4)P Exchange by the ER-Golgi Tether OSBP. *Cell*, *155*(4), 830–843. https://doi.org/10.1016/j.cell.2013.09.056

Moreno, P. R., Purushothaman, M., & Purushothaman, K. R. (2012). Plaque neovascularization: Defense mechanisms, betrayal, or a war in progress. *Annals of the New York Academy of Sciences*, *1254*(1), 7–17. https://doi.org/10.1111/j.1749-6632.2012.06497.x

Niebuhr, K., Giuriato, S., Pedron, T., Philpott, D. J., Gaits, F., Sable, J., … Payrastre, B. (2002). Conversion of PtdIns(4,5)P₂ into PtdIns(5)P by the S.flexneri effector IpgD reorganizes host cell morphology. *The EMBO Journal*, *21*(19), 5069 LP – 5078. https://doi.org/10.1093/emboj/cdf522

Noguera-Troise, I., Daly, C., Papadopoulos, N. J., Coetzee, S., Boland, P., Gale, N. W., … Thurston, G. (2006). Blockade of Dll4 inhibits tumour growth by promoting non-productive angiogenesis. *Nature*, *444*(7122), 1032–1037. https://doi.org/10.1038/nature05355

Olkkonen, V. M. (2015). OSBP-Related Protein Family in Lipid Transport Over Membrane Contact Sites. *Lipid Insights*, *2015*, 1–9. https://doi.org/10.4137/Lpi.s31726

Pico, A. R., Conklin, B. R., Evelo, C. T., Hanspers, K., van Iersel, M. P., Kutmon, M., & Kelder, T. (2011). WikiPathways: building research communities on biological pathways. *Nucleic Acids Research*, *40*(D1), D1301–D1307. https://doi.org/10.1093/nar/gkr1074

Pietrangelo, A., & Ridgway, N. D. (2018). Bridging the molecular and biological functions of the oxysterol-binding protein family. *Cellular and Molecular Life Sciences*, *75*(17), 3079–3098. https://doi.org/10.1007/s00018-018-2795-y

Platre, M. P., & Jaillais, Y. (2016). *Lipid Signaling Protocols* (M. G. Waugh, ed.). https://doi.org/10.1007/978-1-4939-3170-5

Porter, K. R. (1957). STUDIES ON THE ENDOPLASMIC RETICULUM: III. ITS FORM AND DISTRIBUTION IN STRIATED MUSCLE CELLS. *The Journal of Cell Biology*, *3*(2), 269–300. https://doi.org/10.1083/jcb.3.2.269

Potente, M., Gerhardt, H., & Carmeliet, P. (2011). Basic and therapeutic aspects of angiogenesis. *Cell*, *146*(6), 873–887. https://doi.org/10.1016/j.cell.2011.08.039

Prinz, W. A. (2014). Bridging the gap: Membrane contact sites in signaling, metabolism, and organelle dynamics. *The Journal of Cell Biology*, *205*(6), 759–769. https://doi.org/10.1083/jcb.201401126

Pyl, P. T., Anders, S., & Huber, W. (2014). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Raiborg, C., Wenzel, E. M. M., Pedersen, N. M. M., & Stenmark, H. (2016). ER–endosome contact sites in endosome positioning and protrusion outgrowth. *Biochemical Society Transactions*, *44*(2), 441–446. https://doi.org/10.1042/BST20150246

Robertson, J. D. (1960). The molecular structure and contact relationships of cell membranes. *Progress in Biophysics and Molecular Biology*.

Sick, E., Jeanne, A., Schneider, C., Dedieu, S., Takeda, K., & Martiny, L. (2012). CD47 update: a multifaceted actor in the tumour microenvironment of potential therapeutic interest. *British Journal of Pharmacology*, *167*(7), 1415–1430. https://doi.org/10.1111/j.1476-5381.2012.02099.x

Sigoillot, F. D., Lyman, S., Huckins, J. F., Adamson, B., Chung, E., Quattrochi, B., & King, R. W. (2012). A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. *Nature Methods*, *9*, 363. Retrieved from https://doi.org/10.1038/nmeth.1898

Simon, A. (2010). *FastQC: a quality control tool for high throughput sequence data*. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Simon, M. L. A., Platre, M. P., Marquès-Bueno, M. M., Armengot, L., Stanislas, T., Bayle, V., … Jaillais, Y. (2016). A PtdIns(4)P-driven electrostatic field controls cell membrane identity and signalling in plants. *Nature Plants*, *2*(7), 16089. https://doi.org/10.1038/nplants.2016.89

Stenmark, H., & Olkkonen, V. M. (2001). The Rab GTPase family. *Genome Biology*,

2(5), REVIEWS3007. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11387043

Tahergorabi, Z., & Khazaei, M. (2012). A review on angiogenesis and its assays. *Iranian Journal of Basic Medical Sciences*, *15*(6), 1110–1126. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23653839

van Iersel, M. P., Kelder, T., Pico, A. R., Hanspers, K., Coort, S., Conklin, B. R., & Evelo, C. (2008). Presenting and exploring biological pathways with PathVisio. *BMC Bioinformatics*, *9*(1), 399. https://doi.org/10.1186/1471-2105-9-399

van Meer, G., Voelker, D. R., & Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology*, *9*(2), 112–124. https://doi.org/10.1038/nrm2330

Wang, H., Ma, Q., Qi, Y., Dong, J., Du, X., Rae, J., … Yang, H. (2019). ORP2 Delivers Cholesterol to the Plasma Membrane in Exchange for Phosphatidylinositol 4, 5-Bisphosphate (PI(4,5)P2). *Molecular Cell*, *73*(3), 458-473.e7. https://doi.org/10.1016/j.molcel.2018.11.014

Wijdeven, R. H., Jongsma, M. L. M., Neefjes, J., & Berlin, I. (2015). ER contact sites direct late endosome transport. *BioEssays*, *37*(12), 1298–1302. https://doi.org/10.1002/bies.201500095

Xing, C., Lee, S., Kim, W. J., Wang, H., Yang, Y.-G., Ning, M., … Lo, E. H. (2009). Neurovascular effects of CD47 signaling: Promotion of cell death, inflammation, and suppression of angiogenesis in brain endothelial cells in vitro. *Journal of Neuroscience Research*, *87*(11), 2571–2577. https://doi.org/10.1002/jnr.22076

Yilmazel, B., Hu, Y., Sigoillot, F., Smith, J. A., Shamu, C. E., Perrimon, N., & Mohr, S. E. (2014). Online GESS: prediction of miRNA-like off-target effects in large-scale RNAi screen data by seed region analysis. *BMC Bioinformatics*, *15*(1), 192. https://doi.org/10.1186/1471-2105-15-192

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., … Flicek, P. (2018). Ensembl 2018. *Nucleic Acids Research*, *46*(D1), D754–D761. https://doi.org/10.1093/nar/gkx1098

R script used for shRNA off-target prediction

```
library(biomaRt)
library(dplyr)
mart <- useMart("ensembl", dataset = "hsapiens_gene_ensembl")
setwd("/Users/juusotaskinen/Desktop/Gradu/bioinfo/RNAseqstar/")
# Only significantly DE genes are loaded
ORP <-read.delim("ORP_DE_pathway_star/ORPsignificant.txt",
stringsAsFactors=FALSE)
PROT <- read.delim("PROT_DE_pathway_star/PROTsignificant.txt",
stringsAsFactors=FALSE)
# Find only down regulated genes
PROT <- subset(PROT, PROT["log2FoldChange"]< 0)
ORP <- subset(ORP, ORP["log2FoldChange"]< 0)
# Find all the 3 UTR sequences associated with an ENSEMBL ID in each sample
ORPsequences <- getSequence(id = ORP[“ENSEMBL],type=
"ensembl_gene_id",seqType = "3utr",mart = mart)
PROTsequences <- getSequence(id = PROT[“ENSEMBL],type=
"ensembl_gene_id",seqType = "3utr",mart = mart)
# Add seed sequences and 3’UTR sequences to a list
seedORP <- "TTCCCGCC"
seedPROT <- "TTACACCC"
seeds <- list(seedORP,seedPROT)
seq_idsORP <- cbind(ORPsequences[2],ORPsequences[1])
seq_idsPROT <- cbind(PROTsequences[2],PROTsequences[1])
seq_ids <- list(seq_idsORP,seq_idsPROT)
offtargets <- c()
# For loop which iterates over each sample and looks for the seed sequence in
each 3’UTR sequence
for(x in 1:length(seq_ids)){
  seq_id <- seq_ids[[x]] # Take sequence IDs of x sample in the list
  tvect <- grepl(seeds[x],seq_id[,2],fixed = T) # See if seed sequence in
list position x is in any of the 3’UTR sequences, and return FALSE or TRUE
  seq_id <- cbind(seq_id,tvect)
  seqid <- subset(seq_id, seq_id["tvect"] == TRUE) # Store ENSEMBL IDs with
seed sequence in 3’UTR
  offtargets[x] <- seqid #Add ENSEMBL IDs to output list
}
# Make list object into data frame columns to be used in merge()
O <- data.frame(offtargets[1])
colnames(O) <- "ENSEMBL"
P <- data.frame(offtargets[2])
colnames(P) <- "ENSEMBL"
#### Find and save ensembl IDs with seed sequence from RNA-seq data
ORPvsseed <- merge(ORP,O,by.x = "ENSEMBL", by.y = "ENSEMBL" , all = FALSE )
ORPvsseed <- ORPvsseed %>% distinct(symbol, .keep_all = TRUE)
PROTvsseed <- merge(PROT,P,by.x = "ENSEMBL", by.y = "ENSEMBL" , all = FALSE )
PROTvsseed <- PROTvsseed %>% distinct(symbol, .keep_all = TRUE)
write.table(ORPvsseed, file = "ORPseedfoundutr3.txt", sep = "\t",row.names
=TRUE, quote = FALSE)
write.table(PROTvsseed, file = "PROTseedfoundutr3.txt", sep = "\t",row.names
=TRUE, quote = FALSE)
```

GAGE test statistics heatmaps

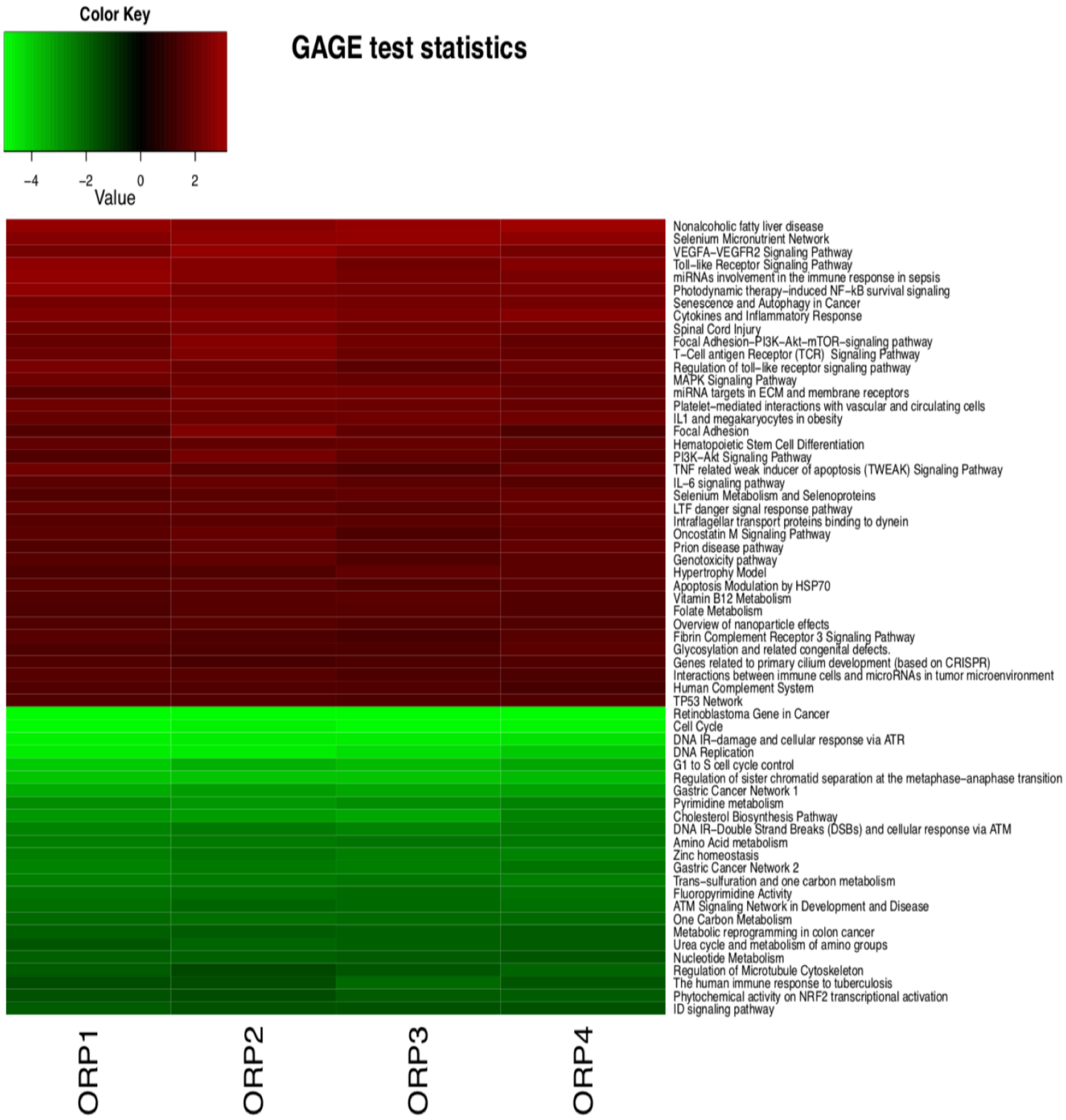


Figure 1.Up- and downregulated pathways in ORP2 silenced samples. Heat map shows the relative change in the pathways compared to control, where green shows the amount of relative downregulation and red shows the relative upregulation.

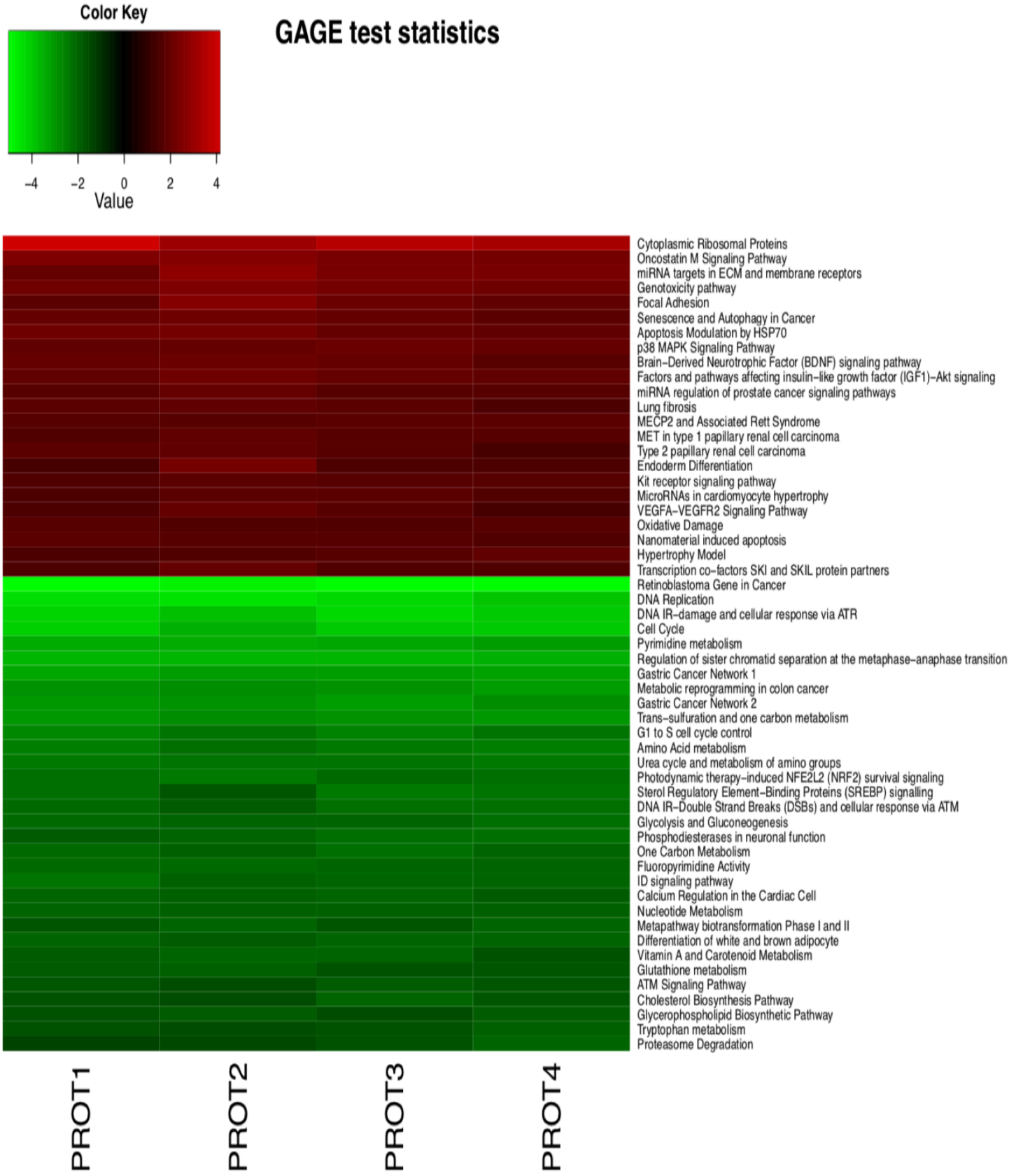


Figure 2. Up- and downregulated pathways in protrudin silenced samples. Heat map shows the relative change in the pathways compared to control, where green shows the amount of relative downregulation and red shows the relative upregulation.